

nature

DOWN TO THE NANOWIRE

A tunable light source for
bio-friendly nanophotonics

END OF THE
BLANK ENA

What will be the
for US science?

CALIFORNIA SMOKING

The ban that
changed the world

PARKINSON'S DISEASE

Rejuvenation therapy
for neurons

All for one...

Many medium-sized university departments feel they are engaged in an unequal struggle against larger and more-entrenched rivals. But there is a way in which they can fight back.

A few years ago, the outlook for physics departments in Scotland was not auspicious. Today, the future is much brighter, thanks to a mechanism that could benefit universities elsewhere. The Scottish Universities Physics Alliance (SUPA), which has forged operational links between six physics departments, looks set to overcome a challenging situation by pooling resources and aiming high.

Like many of their counterparts, the departments in the alliance are fighting for research funds with dozens of rivals, including some elite institutions. They are also competing globally for talented graduate students. Faced with such challenges, medium-sized university departments have few attractive options. They can try to enrol more undergraduates, perhaps by lowering standards, in order to attract block funding that British agencies give to support teaching. In extreme cases, the response has been simply to give up — several chemistry and physics departments have closed in Europe in recent years.

SUPA's founders believe that it is possible to thrive by joining forces. Although this sounds easy, it isn't. Most university department chairs did not work their way up to their positions over decades just to throw whatever influence they have into a pot with five or six of their fiercest local rivals. Universities in every corner of the world experience such rivalries, and variations in quality can make departments reluctant to work with each other on equal terms. And if the universities' top administrators don't buy in, joint ventures between departments are liable to be short-lived.

That may be why the kind of fusion that SUPA has attempted is still relatively rare. But, two-and-a-half years after its formal launch, the experiment is drawing attention from abroad, as well as considerable financial support from both the participating universities and the Scottish Funding Council, which will shortly review a substantial application for funds for the project's second phase.

The alliance is run by a handful of staff in Edinburgh. Physics departments at the participating universities — St Andrews, Edinburgh, Glasgow, Heriot-Watt, Paisley and Strathclyde — have agreed to share authority over hiring decisions and research planning. A board of these departments, together with Ian Halliday, SUPA chief

executive and a Scottish physicist of international standing, meets regularly to steer the alliance. The total operation has some 200 faculty members and 400 graduate students, and attracts about £30 million (US\$60 million) in research grants each year.

At the alliance's annual meeting earlier this month, Halliday said that it is already attracting applications from top-tier candidates for its joint graduate school, and is increasing the calibre of its staff appointments. It remains a challenge to catch the very best students, however, and getting full attendance at video-linked graduate-student classes is also none too easy. Progress in the quality of research — the project's overriding objective — will, of course, take time to achieve.

But if imitation is the highest form of flattery, the project has already enjoyed some measure of success. Similar projects are under way in several disciplines in Scotland and the model has drawn interest from as far

"It is possible to thrive by joining forces — which sounds easy but isn't."

afield as Canada and Spain, where departments face the same basic challenges in trying to compete globally. And university administrators and funding agencies seem to be impressed that departments are ready to do something different in order to succeed.

There is also a political aspect to the alliance's momentum. When Roger McClure, chief executive of the Scottish Funding Council, says that SUPA's strong early run has been "forged by centuries of oppression", the alliance's many English-born physicists enjoy the joke. But he has a serious point: it may be easier for universities in a place such as Scotland to find a genuine common cause than in, say, the midlands of England, where a sense of regional identity is less firmly established.

The SUPA example suggests that, where administrators, department heads and faculty members can find a common cause, it is possible for a group of medium-sized departments to make an impact. Those who draw the appropriate lessons should prosper in a scientific world in which departments must be internationally competitive to survive. ■

Meanings of 'life'

Synthetic biology provides a welcome antidote to chronic vitalism.

Many a technology has at some time or another been deemed an affront to God, but perhaps none invites the accusation as directly as synthetic biology. Only a deity predisposed to cut-and-paste would suffer any serious challenge from genetic engineering as it has been practised in the past. But the efforts to

design living organisms from scratch — either with a wholly artificial genome made by DNA synthesis technology or, more ambitiously, by using non-natural, bespoke molecular machinery — really might seem to justify the suggestion, made recently by the ETC Group, an environmental pressure group based in Ottawa, Canada, that "for the first time, God has competition".

That accusation was levelled at scientists from the J. Craig Venter Institute in Rockville, Maryland, based on the suspicion that they had synthesized an organism with an artificial genome in the laboratory. The suspicion was unfounded, but this feat will surely be achieved in the next few years, judging from the advances reported earlier this

month at the Kavli Futures Symposium in Ilulissat, Greenland, on the convergence of synthetic biology and nanotechnology, and the progress towards artificial cells.

But should such efforts be regarded as 'creating life'? The idea that such creation is a momentous step has deep roots running from the medieval homunculus portrayed by Paracelsus and the golem of Jewish legend to the modern faustian myth of Frankenstein. It will surely be hard to uproot. This is unfortunate, as the idea is close to meaningless.

There is a popular notion that life is something that appears when a clear threshold is crossed. One might have hoped that such perceptions of a need for a qualitative difference between inert and living matter — such vitalism — would have been interred alongside the pre-darwinian belief that organisms are generated spontaneously from decaying matter. Scientists who regard themselves as well beyond such beliefs nevertheless bolster them when they attempt to draw up criteria for what constitutes 'life'. It would be a service to more than synthetic biology if we might now be permitted to dismiss the idea that life is a precise scientific concept.

One of the broader cultural benefits of attempts to make artificial cells is that they force us to confront the contextual contingency of this troublesome idea. The trigger for the ETC Group's protest was a patent filed by the Venter Institute in October 2006 on a "minimal

bacterial genome" — a subset of genes, identified in *Mycoplasma genitalium*, required for the organism to be viable "in a rich bacterial culture medium". That last sounds like a detail, but is in fact essential. The minimal requirements depend on the environment — on what the organism does and doesn't have to synthesize, for example, and what stresses it experiences. Too much minimization and you end up with cells on life support. And participants at the Greenland meeting added the reminder that cells do not live alone, but in colonies and, in general, in ecosystems. Life is not a solitary pursuit, nor can evolution happen without the opportunity for competition.

Synthetic biology's view of life as a molecular process lacking moral thresholds at the level of the cell is a powerful one. And it can and perhaps should be invoked to challenge characterizations of life that are sometimes used to defend religious dogma about the embryo. If this view undermines the notion that a 'divine spark' abruptly gives value to a fertilized egg — recognizing as it does that the formation of a new being is gradual, contingent and precarious — then the role of the term 'life' in that debate might acquire the ambiguity that it has always warranted. ■

"It would be a service to more than synthetic biology if we might now be permitted to dismiss the idea that life is a precise scientific concept."

Electric skies?

How to navigate a flight path to greener air travel.

The widespread availability and affordability of air travel has delivered unprecedented opportunity to travel and a world that is more closely interlinked than ever before. But air transport is also a substantial contributor to greenhouse-gas emissions, leading to a flurry of discussion about what could be done to reduce its 'carbon footprint'.

The airlines and the aerospace industry are increasingly conscious that this concern could put a damper on the growth of their businesses. The Boeing 787, the first exemplar of which is expected to be rolled out of the factory next week and flown next month, is said to be 20% more fuel efficient than the airliners it will replace. Richard Branson's Virgin Airlines said earlier in the year that it plans to begin testing unspecified biofuels in airliners. And EasyJet, a low-cost European airline, has said that it hopes to halve its emissions per passenger kilometre by 2015.

It is by no means clear how much of this is public-relations talk, aimed at deflecting growing public disquiet about the carbon emissions associated with flying. But in the long term, there can be no doubt that the industry will pursue technologies to cut emissions.

One such technology — the use of light carbon-fibre composites in place of aluminium alloys for airframe construction — is incorporated, for the first time in civil aviation history, in the Boeing 787. Further improvements in the strength-to-weight ratio of aircraft structures will come from composites that rely on carbon nanotubes, instead of polymer-based fibres. The cost of the bulk manufacture of

nanotubes is steadily decreasing, although important technical barriers still need to be overcome. Problems associated with optimizing the properties of these composites are discussed in a News and Views Q&A in this issue (see page 1066).

But although the use of new high-performance materials can contribute substantially to the sort of efficiency improvements attained by the 787, truly impressive reductions in airliner emissions would require the industry to take the thoroughly radical (and currently inconceivable) step of replacing the gas turbine engine as the airliners' means of propulsion.

A paper published earlier this month (P. J. Masson *et al.* *Supercond. Sci. Technol.* **20**, 748–756; 2007) puts forward a hypothesis that high-temperature superconducting technologies could provide a route to all-electric aircraft that burn hydrogen fuels and have electrical systems powered by fuel cells. Although highly speculative as to its practicability, it would greatly reduce carbon emissions assuming also that the hydrogen could be produced by nuclear or renewable technologies.

If 'clean' air travel is indeed achievable, it remains a long way off, and will only be developed through a concerted effort between scientists, engineers, governments and businesses. At the Paris Air Show last week, European research commissioner Janez Potočnik announced a 'clean sky' research initiative, under which the European Commission will invest €800 million (US\$1 billion) from 2008, while hoping to attract a similar amount from private industry. But it will take more than that. Equally commendably, Louis Gallois, the chief executive of Airbus, called for an unprecedented meeting of airliner and aeroengine makers (including Airbus and arch-rival Boeing) later this year to discuss global collaboration on the technical challenges ahead. These obstacles are considerable, but such an approach gives hope of surmounting them. ■

RESEARCH HIGHLIGHTS

ORGANIC CHEMISTRY

Ring leaders

Angew. Chem. Int. Edn doi:10.1002/anie.200701551 (2007)

Chemists in the United States report an improved procedure for preparing boronate esters — starting materials for making the many molecules that contain linked aromatic rings.

The preferred route to boronate esters is from readily available aryl chlorides. But existing processes require high temperatures and don't work for all substrates. By changing the catalysts used, Stephen Buchwald and his colleagues at the Massachusetts Institute of Technology, Cambridge, prepared boronate esters from aryl chlorides at room temperature. Their method also works for a wider range of molecules than the previous method.

Finally, they react two aryl chlorides together, through a boronate ester intermediate, to create 'biaryl' compounds. This is the first time that this desirable transformation has been achieved directly from two aryl chlorides.

CHEMICAL BIOLOGY

Smart drops

J. Am. Chem. Soc. doi:10.1021/ja072292a (2007)

Artificial cell-like compartments can be linked into networks that act as devices, show Matthew Holden of the University of Oxford and his co-workers.

Each compartment is a water droplet, typically less than 1 millimetre across, surrounded by a lipid monolayer and immersed in oil (pictured below). When two droplets stick together, they are separated by a cell-membrane-like lipid bilayer, which can host membrane proteins such as ion channels that allow communication between the 'cells'.

The researchers created a chain of three droplets containing solutions of different ions with channels in the connecting walls to pump the ions. This acted as a 'biobattery',



Assorted fungi

Mycol. Res. 3, 509–547 (2007)

A concerted effort by 67 researchers based in 13 countries to sort out the taxonomy of fungi has reached its conclusion. The hope is that the new classification, which divides the kingdom Fungi into 195 taxa, will clear up naming confusion in the literature and bring about consistency across online databases.

To devise the classification, David Hibbett of Clark University in Worcester, Massachusetts, and his colleagues sorted through molecular data on different fungal species and examined the history of their nomenclature. Their effort, part of the 'Assembling the Fungal Tree of Life' project, should provoke some changes. For example, the researchers propose that the phylum Zygomycota, which includes the 'sugar' moulds found on fruit, is disbanded and its taxa shared among other groups.



D. HIBBETT

generating a current. Also, a network of droplets connected through the light-sensitive proton pump bacteriorhodopsin offered an electrochemical light-meter that mimics light-detecting retinal cells.

MEDICINE

Brain boost

Lancet 369, 2097–2105 (2007)

A preliminary gene-therapy trial has improved brain function in patients with Parkinson's disease.

Parkinson's disease causes neuronal degeneration, loss of motor control, and reduced levels of an important neurotransmitter known as GABA. Matthew During of Ohio State University in Columbus and his colleagues placed a gene important for GABA production, called glutamic acid decarboxylase, into a virus. They then injected the virus into a specific region of the brain, on one side only, in 12 patients with Parkinson's disease.

The phase I trial was designed to test safety rather than efficacy, but brain scans and motor-control tests showed that function in the injected side of the brain improved relative to the side that had not received the gene. No patients died or developed new neurological deficits.

CANCER BIOLOGY

To drug the undruggable

EMBO J. doi:10.1038/sj.emboj.7601744 (2007)

An antibody fragment can hit a cancer target that many have been deemed 'undruggable', say scientists in the United Kingdom.

Mutations that activate Ras proteins show up in as many as 30% of cancers, but the difficulty of blocking their protein–protein interactions inside the cell had made them seem intractable targets.

Terence Rabbitts, now at the Leeds Institute of Molecular Medicine, and his colleagues report that an antibody fragment, dubbed iDab#6, jams mutant Ras by blocking a key interaction site. In mice injected with human tumour cells, tumour growth stopped if the cells expressed the antibody fragment.

Delivering the genetic material to express iDab#6 in human patients would be a challenge, but the team's characterization of the Ras–antibody interaction may also help small-molecule drug development.

CELL BIOLOGY

Numbing the pain

Neuron 54, 905–918 (2007)

Intense pain can switch off the ion channels that sense it by flipping a molecular toggle,

researchers report. The ion channel TRPV1 regulates the flow of ions such as calcium into nerve cells in response to heat and acidity, mediating the feeling of burning pain. Rachele Gaudet and her colleagues at Harvard University in Cambridge, Massachusetts, determined the structure of a TRPV1 region that protrudes into the cell, known as the ankyrin-repeat domain. They found that it binds to one or other of two signalling molecules: ATP or the calcium sensor calmodulin.

The channel responded to repeated stimulation when bound to ATP, but its response was deadened by calmodulin. The team proposes that the high intracellular calcium levels that follow channel stimulation favour calmodulin binding and thus channel desensitization.

BIOCHEMISTRY

A new last resort

J. Am. Chem. Soc. doi:10.1021/ja068602r (2007)

The arms race between antibiotics and bacteria has yielded bugs that are impervious to vancomycin, the antibiotic 'of last resort'. Now, researchers have found a simple way to create vancomycin-like compounds that are up to 40 times more potent than their predecessor.

Vancomycin comprises a loop of three amino-acids with two attached sugars. Jon Thorson of the University of Wisconsin-Madison and his colleagues produced eight vancomycin variants by adding various glucose molecules containing 'lipid-like' hydrocarbons to vancomycin's amino-acid backbone.

Biological activity against vancomycin-resistant bacteria was highest when the hydrocarbon was attached at the third or fourth carbon of the glucose molecule. The method provides a way of generating libraries to optimize antibiotics for targeting resistant pathogens.

CELL BIOLOGY

Act in two parts

Science **316**, 1749–1752 (2007)

Researchers have unpicked a mechanism of actin's dual function in cells.

Actin is a major component of a cell's cytoskeleton and a regulator of gene

expression. These functions are linked through a protein known as MAL that binds actin and that normally shuttles between a cell's nucleus and cytoplasm.

When a cell is preparing to divide, actin assembles into cytoskeletal polymers. The resulting drop in free actin levels causes MAL to accumulate in the nucleus, where it promotes the expression of growth-related genes.

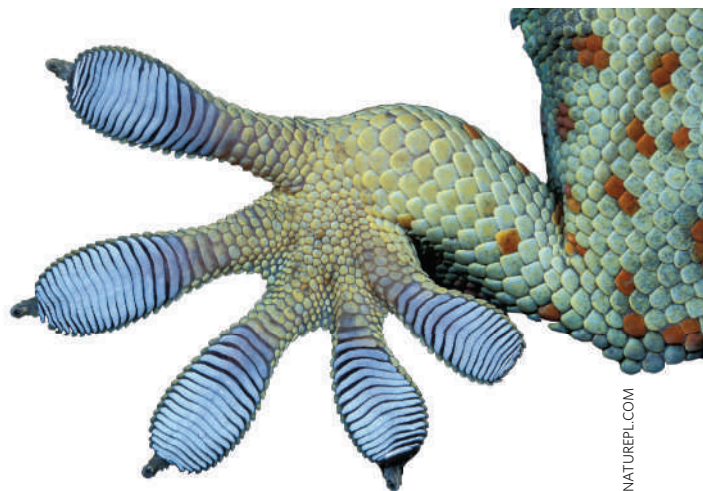
Richard Treisman and his colleagues at Cancer Research UK's London Research Institute show that MAL accumulates in the nucleus because its export to the cytoplasm requires it to be bound to nuclear actin. The team also shows that actin binding to MAL in the nucleus inhibits gene expression.

MATERIALS SCIENCE

Geckos outstuck

Proc. Natl Acad. Sci. USA **104**, 10792–10795 (2007)

A reusable adhesive tape four times stickier than gecko feet has been made by Ali Dhinojwala at the University of Akron, Ohio, and his colleagues.



L. ARNDT/NATUREPL.COM

Gecko feet stick to surfaces using pads covered with rod-like protrusions called setae, which split at their end into smaller filaments known as spatulas, just a few nanometres thick. These fine hairs penetrate every crevice on a surface, then stick via van der Waals forces.

Carbon nanotubes, which are strong and stiff, have previously been identified as candidate hairs for a synthetic gecko pad; now Dhinojwala and his team have marshalled them into pillar-like clusters with square cross-section, copying the hierarchical organization of setae. One square centimetre of tape covered with these structures can support nearly 4 kilograms.

JOURNAL CLUB

Achim Müller

University of Bielefeld, Germany

A chemist finds beauty in molecules that resemble an early model of the Solar System.

Since Plato's time, people have been fascinated by the beauty of highly symmetrical objects. The symmetry of the C₆₀ buckyball surely contributed to scientists' tremendous interest in this spherical molecule. Indeed, I was convinced that the discovery of C₆₀ would induce a rush among chemists to search for other symmetrical structures.

That rush may not have happened, but scientists have still turned up some surprising highly symmetrical structures. A recent report from researchers at Xiamen University in China (X.-J. Kong *et al. J. Am. Chem. Soc.* **129**, 7016–7017; 2007) describes a cluster in which beauty cages beauty; it consists of an icosidodecahedron of nickel ions, having 20 triangular faces and 12 pentagonal faces, inside of which sits a dodecahedron of lanthanum ions.

The team describes the magnificent structure as 'Keplerate', a term that I and my colleagues first used around ten years ago to describe structures that contain Platonic and Archimedean solids (regular polyhedra, and polyhedra with two types of face, respectively) one inside another, like Russian dolls. It honours Johannes Kepler, who in the sixteenth century developed a model of the cosmos in which "the radii of the successive planetary orbits are proportional to the radii of spheres that are successively circumscribed around and inscribed within the five Platonic solids".

Another recent report found these same shapes — the icosidodecahedron and dodecahedron — in Keplerate-type arrangements in quasicrystals (H. Takakura *et al. Nature Mater.* **6**, 58–63; 2007). Such crystals are still poorly understood. I hope that future work will correlate these materials' properties with their beauty.

Discuss this paper at <http://blogs.nature.com/nature/journalclub>

NEWS

The techniques used to store tissue samples might distort the genetic information.



H. MORGAN/SPL

Cancer atlas maps out sample worries

America's National Cancer Institute (NCI) has confirmed that its much-heralded cancer genome atlas is off to a slow start. And scientists say that the obstacles facing the project are just a taste of those that will soon confront all researchers who use large-scale methods to analyse material from tissue banks.

Nine months ago, US National Institutes of Health director Elias Zerhouni announced the "scientific start" of the Cancer Genome Atlas, which aims to catalogue the genetic characteristics of various cancers at a previously impossible level of detail. The project was controversial even before it started, with some scientists afraid that it would divert resources from investigator-initiated research in tight budget times (see *Nature* 438, 894; 2005). But it prevailed, thanks in part to strong support from Francis Collins, director of the National Human Genome Research Institute (NHGRI) in Bethesda, Maryland.

Now, the NCI has confirmed that the project stumbled as soon as it left the starting blocks. The problem lies with brain-cancer samples, which are removed surgically from patients and stored at a tumour bank at the MD Anderson Cancer Center in Houston, Texas. The centre was chosen to provide the Cancer Genome Atlas with samples of glioblastoma multiforme

— a type of brain cancer that is almost always fatal. Glioblastoma is one of three cancers that will be analysed in the atlas's US\$100-million three-year pilot programme, which is funded jointly by the NCI and the NHGRI. The other two diseases being analysed are lung and ovarian cancers.

The pilot aims to analyse at least 500 samples of each cancer. But this February, when scientists started to retrieve tissue samples from the bank, "we started realizing the scene wasn't as rosy as we had hoped", says Carolyn Compton, director of biorepositories and biospecimen research at the NCI.

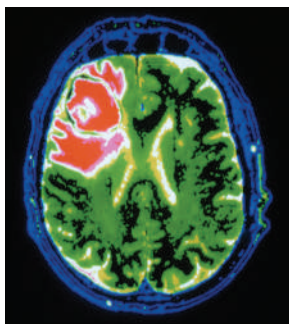
So far, the project has analysed fewer than

half the number expected because about half of the samples at the MD Anderson bank haven't met the project's stringent quality tests, according to cancer geneticist Ronald DePinho from Harvard Medical School in Boston, Massachusetts, who is co-chair of the atlas's external scientific committee.

Compton and other scientists say this problem isn't unique to the MD Anderson tumour bank. Variable sample quality is a concern with all tissue banks, scientists say, because every bank has its own standards, and samples aren't usually checked for quality until someone needs them for a project.

"We haven't had the large number of samples we expected," says Lynda Chin, a cancer geneticist at the Dana-Farber Cancer Institute in Boston, Massachusetts, and co-head of one of the atlas's seven Cancer Genome Characterization Centers. "But this is the reality of the limitations of the existing tissue banks."

The Cancer Genome Atlas set high standards for its samples to ensure high-quality data. All the samples included in the project must have at least 80% viable cells, because the project is analysing DNA and RNA only from live cells. The samples also have to be quite large — weighing at least 200 milligrams —



Deadly diseases: the fatal brain cancer glioblastoma can be hard to sample.

PHOTOTAKE INC./ALAMY



OLDER SIBLINGS ARE SMARTER

Social standing within a family affects average intelligence scores.

www.nature.com/news

CORBIS

NASA/SPL

because the tissue has to be shared between the many analysis centres involved in the project.

The high standards are part of the reason that the project stumbled with the glioblastoma samples. Glioblastoma tumours tend to be spread throughout the brain, and must by definition contain necrotic, or dead, tissue. So scientists expected that some tumours from the MD Anderson bank would contain too much necrotic tissue to meet their criteria — but not this many.

The failure of so many samples has prompted the Cancer Genome Atlas to expand its criteria to admit brain-cancer samples with more necrotic tissue. It has also started to take samples from brain tumour banks at the Henry Ford Health System in Detroit, Michigan, and the University of California, San Francisco.

Compton says that the lung and ovarian cancer projects shouldn't have the same problems as the one for glioblastoma, because those tumours tend to be larger. She still hopes that the pilot project will complete its work in the scheduled three years.

But changing the criteria for glioblastoma samples won't solve the bigger problems with tissue banking. Foremost among them is how the procedures used in tissue banking affect the nucleic acids and proteins in the stored samples. For instance, it's not clear how anaesthesia affects gene expression during surgery to remove tumours, Compton says. That could mean that signatures gleaned from banked tissues might just be by-products of procedures used in surgery and storage — not markers of disease processes. "The state of the science right now is pathetic," Compton says.

Because the science is so poor, the NCI has decided, for the first time, to fund studies on biospecimen banking. On 28 June, Compton is scheduled to present the agency's biospecimen research plan to its board of scientific advisers. If it is approved, the NCI will award scientists money to study how collection and banking processes affect the nucleic acids and proteins embedded within tissue samples.

The NCI is also working with research agencies around the world to try to encourage tissue banks everywhere to adopt uniform standards. Researchers in many countries are grappling with the same problem. The International Agency for Research on Cancer, a World Health Organization body based in Lyon, France, with 20 member states, is also trying to address it.

In the meantime, DePinho insists that the Cancer Genome Atlas is doing well for a new venture of its scope. "This is an extraordinarily difficult project, so these hiccoughs at the embryonic stage are to be expected," he says. ■
Erika Check

Geophysicists combine forces

An ambitious programme aimed at freeing up Earth-science data for researchers and educators worldwide will be launched next week at the quadrennial meeting of the International Union of Geodesy and Geophysics in Perugia, Italy. It is one of several efforts by geoscientists to improve access to geophysical data on the web.

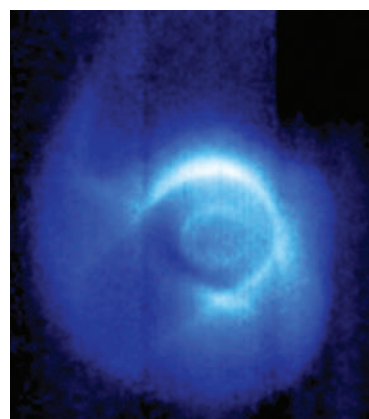
Called the Electronic Geophysical Year (eGY), the programme has been organized by a broad range of researchers and organizations and will promote the creation of virtual observatories to make geophysical data easier to find and use. It takes its inspiration from the fiftieth anniversary of the 1957 International Geophysical Year, which was the first concerted effort to understand the Earth as a physical system.

Sharing and transferring data is much easier than it was during the original International Geophysical Year. But because different research groups often have data stored in incompatible formats, researchers can't always combine the various sources for study. Efforts to transfer historical data languishing in archives from deteriorating paper to accessible digital formats are piecemeal and require prioritization.

One of eGY's goals is to overcome such hurdles, in part through the use of virtual observatories that will serve as clearinghouses for scientific collections. "This is the forefront of information technology: to standardize computer-language syntax," says William Peterson, a physicist at the University

of Colorado in Boulder and an eGY organizer. "We hope to build layers of software so anyone can display and integrate data in different forms." On 12–15 June, around 75 researchers from around the world attended the Virtual Observatories in Geosciences conference in Denver, Colorado, to plot this course.

A case in point, says Jeffrey Love, a geomagnetist at the US Geological Survey office in Denver, are records of Earth's magnetic field that extend



Force field: Earth's magnetosphere could shed light on magnetic storms.

back to the nineteenth century. Decades-old data originally gathered to help marine navigation at labs in Maryland, Hawaii and Alaska are now of interest to scientists studying magnetic storms in Earth's magnetosphere that occur as a result of solar activity. "It is a daunting task," says Love, "making sure people know what they have and should save."

Other international efforts are aimed at collecting new data. One part of the current International Heliophysical Year is a concerted effort to extend low-cost data-collection systems to developing nations. In Africa, for instance, 35 nations have agreed to the installation of

special Global Positioning System pods. These use the signals from navigation satellites to measure electron density in the ionosphere (where the atmosphere meets space). Such monitoring is widespread in the United States, Europe and Latin America, but data for Africa are weak, says Joseph Davila, a physicist at NASA's Goddard Space Flight Center in Greenbelt, Maryland, who is leading the effort. Last week, Davila was in Libya helping to

install a radio receiver at the University of Sheba, south of Tripoli, as part of AWESOME (Atmospheric Weather Electromagnetic System for Observation, Modeling and Education), another ionospheric monitoring project.

In 2008, the United Nations-sponsored Year of Planet Earth will be looking at similar efforts to further its goal of promoting "Earth Sciences for Society". Eduardo de Mulder of

the Geological Survey of the Netherlands in Utrecht, a leader on the Planet Earth project, says the organization is trying to raise at least US\$5 million to fund international efforts that would include 60 nations. Scientists are applying for grants from this pool of funds, with the first awards expected late this year.

Among the project's main efforts will be an annotated geological map of the world, onegeology.com, that would provide access to all sorts of data. "It would be a living process," says de Mulder, "with the public and governments using the Earth science information to make better policy decisions." ■

Rex Dalton

GRAPHIC DETAIL

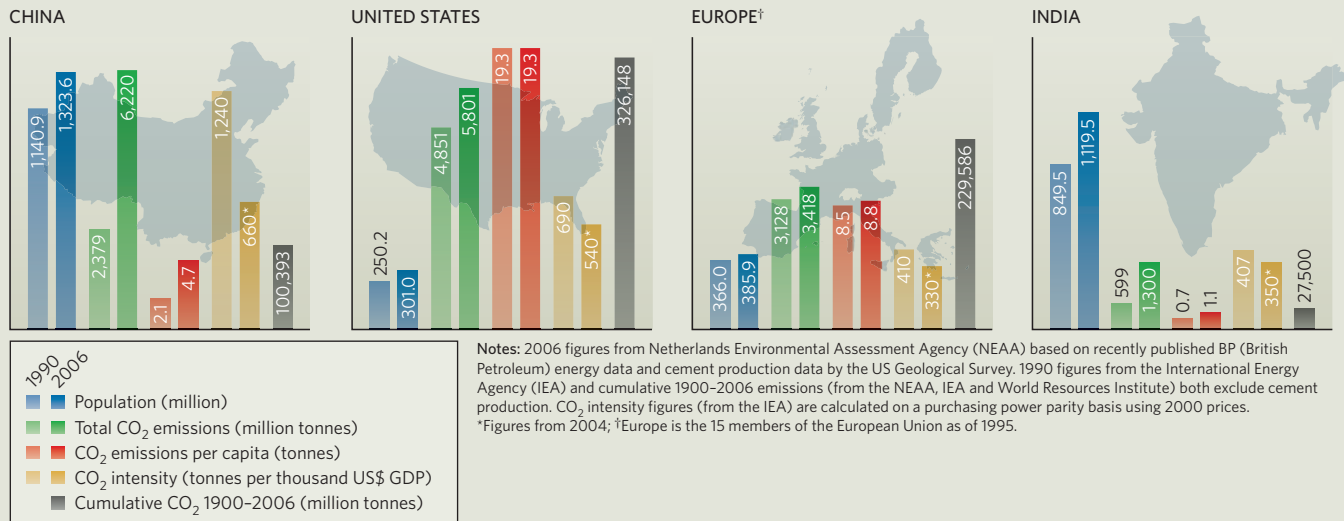
Gas exchange: CO₂ emissions 1990–2006

Last week, the Netherlands Environmental Assessment Agency produced a preliminary report showing that China had overtaken the United States as the world's largest emitter of carbon dioxide from the burning of fossil fuels and the manufacture of cement (44% of the world's new cement is currently being laid in China).

Here's how the world's big emitters stacked up.

In per capita terms, the United States is still easily the most carbon-profligate economy, and it has made by far the largest historical contribution to the stock of atmospheric CO₂. In terms of the emissions it takes to provide a given amount of gross domestic product

(GDP), the carbon intensity, China is in the worst position. The carbon intensity has dropped in all four economies since 1990, most impressively in China. But given economic growth, overall global CO₂ emissions rose by more than 35% between 1990 and 2006.



Darfur's climate roots challenged

A United Nations (UN) report claiming that “climate change, land degradation and the resulting competition over scarce natural resources are among the root causes” of the Darfur conflict has been met with scepticism by experts on the region. Although these factors contributed, they say, the UN overstates the case.

The 358-page Sudan Post-Conflict Environmental Assessment, released on 22 June by the UN Environmental Programme, is in the main a comprehensive treatise on environmental management. “It seems to have been thoughtfully and professionally done,” says Martha Saavedra, associate director of the Center for African Studies at the University of California, Berkeley.

But like many experts, Saavedra questions the UN's spin, which has played up the role of environmental degradation. Ban Ki-moon, UN Secretary-General, stressed this theme in a 16 June article in the *Washington Post* entitled “A Climate Culprit In Darfur”. His piece painted a tragic picture of resource scarcity triggering conflict between herders and settled farmers who had previously coexisted peacefully.

“Almost invariably, we discuss Darfur in a convenient military and political shorthand — an ethnic conflict pitting Arab militias against black rebels and farmers,” Ban wrote. “Look

to its roots, though, and you discover a more complex dynamic. Amid the diverse social and political causes, the Darfur conflict began as an ecological crisis, arising at least in part from climate change.”

Alex de Waal, an expert on Darfur at the Social Science Research Council in New York, agrees that the dynamic is complex — and warns that the environmental argument, too, raises “a danger of oversimplifying Darfur.” Darfurians have adapted to environmental change for centuries, he points out. “Over the past thirty



Do Darfur's problems stem from resource scarcity?

years, change has occurred at a faster pace and on a larger scale,” he says. “But depleted natural resources and livelihood transformations cannot on their own account for armed conflict.”

The true culprit in Darfur is the National Islamic Front, which came to power in Khartoum in a military coup in 1989. It was bent on expanding its political base through ethnic cleansing, using terror as a tactic, says Eric Reeves, an analyst of Sudan at Smith College in Northampton, Massachusetts. “This is not about competition over resources.”

“There will always be conflict, but having good, regular conflict regulating and resolution mechanisms — legitimate governing systems — is key,” agrees Saavedra, adding that although these issues are referred to in the report, the political issues that lie at the heart of the conflict are sidestepped.

A version of the Darfur conflict would have occurred whatever the environmental changes, says Reeves, adding that they cannot be used as an alibi for the international community's failure to prevent the genocide. “You don't have a diminishment in rain, and then an uncontrolled genocidal conflict; this just doesn't correspond to the political realities in Darfur.”

Declan Butler

M. KAMBER/REUTERS

Treaty caution on plankton plans

The scientific advisory group for the parties of the London Convention, the main treaty governing ocean dumping, has for the first time questioned the use of 'iron fertilization' in combating climate change. The group's statement, released on 22 June, will not in itself interfere with plans by Planktos, a San Francisco-based company, to create a large plankton bloom with the technique in the coming months. But this official statement of concern could strengthen calls for further regulation under the terms of the treaty when the partners meet later this year.

The London Convention advisory group, which includes representatives from 19 parties to the convention, as well as various observers, says in its public statement that "knowledge about the effectiveness and potential environmental impacts of ocean iron fertilization currently is insufficient to justify large-scale operations".

Although there has been research into iron fertilization over the past decade, just how much carbon such blooms sequester is still unclear. A key complicating factor is that much of the resulting biomass may be converted back to carbon dioxide through grazing by other organisms. Scientists

and environmentalists have also raised concerns about potential side effects, including the growth of harmful algal species and the depletion of oxygen as a result of bacterial breakdown of the phytoplankton.

Planktos's plan to seed an area roughly 560 kilometres west of the Galapagos Islands would make it the first commercial company to conduct large-scale iron fertilization, although other for-profit groups are also pursuing the idea. Planktos's president, Russ George, says the experiment is designed in large part to answer the very questions raised by the advisory group.

Others disagree. "I think selling [the Planktos project] as research is really quite disingenuous," says David Santillo of the Greenpeace Research Laboratories in Exeter, UK. Greenpeace has campaigned against commercial iron fertilization, and George believes such efforts led to the advisory group taking up the issue.

Planktos plans to release 90 tonnes of haematite, an iron-containing mineral that, although readily available as an industrial waste product, has been passed over in scientific

experiments in favour of more reactive forms of the metal. The company intends to study the seeded area during the six months the bloom is expected to persist, and Planktos says it wants to involve outside scientists in the project. However, according to George, "countless false and harmful statements being presented in the world media" about the company's work have made such collaborators hard to find.

David Kubiak, a spokesman for Planktos, says that, in addition to gathering data, the company hopes to sell 'carbon credits' based on the Galapagos work, subject to outside verification that carbon really is being stored and not re-released. He adds that if between 2.5 million and 4.5 million tonnes of carbon were sequestered, and buyers could be found, the company might make up to \$25 million.

Ken Johnson, a leader of past iron-enrichment research who works at the Monterey Bay Aquarium Research Institute in California, says the estimate is in line with the more optimistic estimates for ratios of bloom carbon to iron, but that other estimates may be two orders of magnitude lower. "The gist of it is we don't know if it's viable or not."

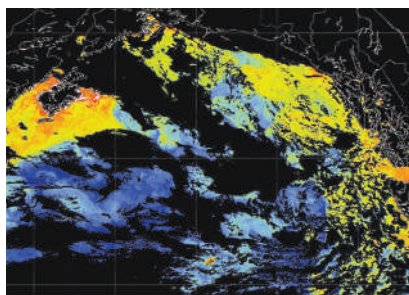
Although Johnson is adamant that such work

should be done by governments, not commercial organizations, he thinks iron fertilization should remain an option for controlling greenhouse gases if it can be made to work. "It's probably not good," he says, "but is it worse than doing nothing at all?"

With the issue of iron fertilization now officially raised by the advisory group's statement, the parties to the London Convention may opt to clarify related legal issues at their next meeting in London in November. Their options would include amending the convention or for signatory countries to amend their respective ocean dumping laws.

Concern in the United States that the Galapagos experiment might use a non-US-flagged ship and thus avoid US regulation was what led the United States to request that the London Convention scientific group consider the issue. Dale Kemery, a spokesman for the US Environmental Protection Agency, says that, regardless of the ship used, the United States might still have jurisdiction if the iron to be released originated there.

Mark Schroppe



Experiments that use iron to seed plankton blooms (yellow) are visible from space.

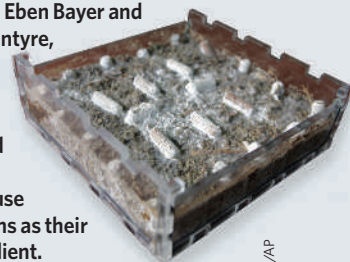
SEAWIFS PROJECT/NASA

SCORECARD



Mushrooms

Fungi may soon become the latest addition to the menu for eco-builders, thanks to Rensselaer Polytechnic graduates Eben Bayer and Gavin McIntyre, who have invented organic household insulation tiles that use mushrooms as their key ingredient.



M. GROLL/AP



BlackBerrys

French defence experts have warned senior government staff not to use the hand-held e-mail devices for official business, citing fears that the messages, which pass through servers in Britain and the United States, are not safe from spies.



S. MAY/ALAMY

ON THE RECORD

"The work is quite unpleasant. We have to stay in a lab smelling those awful gases repeatedly."

Liu Jingcai, one of China's new 'professional noses', who have been trained to sniff out dangerous pollution in the southern industrial town of Panyu.

NUMBER CRUNCH

103 is the average IQ of adults who were born first among their siblings, according to a Norwegian study.

100 is the average IQ of second-born children, showing that older siblings generally grow up to be the cleverest.

4 is the number of older siblings of Charles 'father of biology' Darwin, proving that there's no accounting for genius.

Sources: Associated Press, Reuters, Science

SIDELINES

Circumcision for HIV needs follow-up

The French National AIDS Council (CNS), an independent government advisory body on AIDS issues, last week urged prudence in implementing male-circumcision programmes to reduce the spread of HIV. It cautioned that over-zealous roll out could lead to a false sense of security and exacerbate the problem.

In March, the World Health Organization (WHO) endorsed the promotion of male circumcision as a tool for HIV prevention, alongside condoms and antiretroviral drugs. The move was based on a WHO expert consultation, which concluded that the evidence for the intervention's efficacy was "compelling". Recent criticism of this view has been received with anger (see 'Cutting criticism').

The strongest evidence comes from three recent studies. In 2005, a study¹ of 3,300 heterosexual men living in and around Orange Farm, South Africa, by France's National AIDS Research Agency, showed that circumcising men reduced the risk of infection by 60%. Similar levels of protection were found in 2007 by American-funded studies in Kisumu², Kenya, and Rakai³, Uganda.

Taken together with results from observational studies, this is "as convincing evidence as one ever gets in public health," says Helen

Weiss, a statistical epidemiologist at the London School of Hygiene & Tropical Medicine.

But questions remain about the intervention's applicability as a preventative tool on a large scale, says Willy Rozenbaum of the Pierre and Marie Curie University in Paris, who led the CNS study. Rozenbaum questions the speed with which the WHO has acted, and thinks the organization's endorsement — although detailing the caveats — has been misunderstood

in some quarters as saying circumcision is a "miracle solution".

Rozenbaum notes that the effect of circumcision on HIV prevalence in a population will depend on a host of social and cultural factors. There is a real risk, he says, that after circumcision men may have a false sense of security and increase their number of partners, or dispense with condoms. The report also points out that circumcision leaves men more vulnerable to



Circumcision reduces the spread of HIV, but needs to be combined with public education.

Cutting criticism

A controversial recent publication argues that the evidence from cross-country statistics does not support male circumcision as a key intervention against HIV. But according to several angry AIDS researchers, the paper merely shows that the peer-review system of the journal that published it, *PLoS One*, failed on this occasion. The study, they say, is flawed and, moreover, concerns a debate over statistical techniques that, in this instance, have been largely superseded by more powerful clinical trials.

"Size matters: the number of prostitutes and the global HIV/AIDS pandemic," by John Talbott⁴, a former investment banker, was published on 20 June. In a comparison of HIV prevalence across 77 countries, it argues that the effect of male circumcision is overstated, and that the prevalence

of prostitution is the dominant factor. Among other things, Talbott takes issue with the way that some previous studies have 'weighted' circumcision figures according to the size of national populations.

Ecological analyses such as Talbott's, which look for correlations between HIV prevalence and various factors, are useful for generating hypotheses, but don't prove anything, says Anne Buvé, an expert in HIV prevention at the Institute of Tropical Medicine in Antwerp, Belgium. In her opinion, the paper consists of "sweeping generalizations that add nothing substantial to our state of knowledge".

In understanding the effect of male circumcision "we are past this sort of ecological research", says Buvé, pointing to recent randomized clinical trials showing a major protective effect. Talbott's

paper mentions one such study in passing. Talbott declined to comment for publication.

"Our data were never intended to be used in this way," says Buvé, a co-author on the 2006 paper from which Talbott took the preliminary data that he draws many of his

"The paper adds nothing substantial to our state of knowledge."

figures on prostitution from⁵. A host of factors leads to under- and overestimates of the number of prostitutes nationally. There are also problems in using the Muslim proportion of the population as a surrogate for the level of circumcision, experts say.

Nonetheless, the press release made available by *PLoS One*

describes the paper as containing "important findings that should impact policy decisions". It recommends "mandatory" testing and treatment of prostitutes, adding that opposition to such control methods is comparable to "claiming that a rabid dog must be allowed to run free in a neighborhood regardless of how many men women and children he infected and killed".

These are Talbott's words. It is *PLoS One*'s normal practice to publicize any text supplied by the author as a press release, adding a brief disclaimer at the end.

Talbott's paper is a baptism of fire for the young journal, launched in August 2006. The journal is testing a new publishing model, whereby papers are published quickly after being assessed by peer reviewers on technical soundness alone. The idea is that the papers' reliability

infection if they don't refrain from having sex until the wounds have fully healed and that women may find it more difficult to insist on condom use with circumcised men.

Given these concerns, the CNS calls for implementation schemes to be accompanied by public education as well as research that considers the behavioural and cultural factors that affect HIV transmission. Such research is already under

"The impact of circumcision on HIV prevalence will depend on a host of social and cultural factors."

way at various sites. At Orange Farm, a 5-year follow-up study of 30,000 people will look at how the implementation of a real circumcision campaign affects risk behaviours, and

overall HIV prevalence in the population.

Tim Farley, an official in the WHO's HIV prevention team in Geneva, says he "absolutely agrees" with the CNS view, and the need for such research. "These are concerns that need to be studied, but they are not a reason to stop," he says. "We must move ahead in the knowledge that there could be riskier behaviour."

Declan Butler and Lucy Odling-Smee

1. Auvert, B. *et al.* *PLoS Med.* **2**, e298 (2005).
2. Bailey, R. C. *et al.* *Lancet* **369**, 643–656 (2007).
3. Gray, R. H. *et al.* *Lancet* **369**, 657–666 (2007).
4. Talbot, J. R. *PLoS One* **2**, e543 (2007).
5. Vandepitte, J. *et al.* *Sex. Transm. Infect.* **82** (Suppl. 3), iii18–iii25 (2006).

and significance will be further ascertained by online comments and discussion by other scientists that are posted on *PLoS One*'s website after publication.

"The paper is total drivel, it should have been picked up in the review process," claims Tim Farley, an official in the World Health Organization's HIV-prevention team in Geneva. "And coming from *PLoS One* gives [the views in the press release] a public perception of validity. In public health there are severe dangers in such stuff getting through."

"There are lessons to be learned from all papers that we publish; we are a young journal," says Chris Surridge, *PLoS One*'s managing editor, adding that its peer-review model is constantly under refinement. "We are feeling our way."

In this case, the paper was reviewed by a member of *PLoS One*'s editorial board and one external referee. The review criteria of *PLoS One* are that "the data are sufficient to be published and the conclusions not radically overstated by the data," Surridge explains. "I hope we get some discussion of this paper [in the online post-publication comments]."

D.B. and L.O.-S.



Plan for bomb design falters

Congress is losing its appetite for the Reliable Replacement Warhead (RRW), a nuclear bomb conceived as a way to allow weapons physicists to ply their trade in a weapons-ban world. Instead, Capitol Hill is calling for a full review of the US nuclear programme.

After its 4 July recess, the House of Representatives will vote on a bill that would eliminate funding for the RRW from the 2008 budget. The Bush administration had requested \$89 million for the project. And committees in both the House and the Senate are drafting legislation that would limit work on the new design to feasibility and cost assessments. That legislation would also call for a review of the need for nuclear weapons.

Until last autumn, the RRW programme enjoyed bipartisan, if tentative, support in Congress. The project, which aims to design a weapon more robust than its predecessors but that requires no testing (see *Nature* 442, 18–21; 2006), was seen by congressional appropriators as a way to cut the cost of maintaining the nuclear stockpile. Some also liked the idea that it could reduce the size of the stockpile.

But a costly plan released in December on the future of the nuclear weapons complex suggested it would be hard to realize any savings from the RRW. And a study showing that existing warheads will last for at least another 50 years (see *Nature* 444, 660–661; 2006) made the new bomb seem less necessary. Billions have been spent on 'stockpile stewardship' to prolong the lives of current warheads, Congressman Pete Visclosky (Democrat, Indiana) said on 19 June. Now, he said, "We are told: 'Let's do something else'."

"I don't think it is asking too much for a comprehensive nuclear strategy before we build a new nuclear weapon," added Visclosky, who chairs the subcommittee that withdrew the RRW funding.

A separate defence bill working its way through Congress might provide such a strategy. Versions of the authorization bill, which would set out policy but not funding, call for a comprehensive review of US nuclear policy. This would assess whether the size of the stockpile is appropriate in the post-cold-war era and whether an RRW would be necessary for a credible deterrent.

Arms-control advocates are elated by the move. "What Congress has done is really stunning," says Philip Coyle, a senior adviser at the Center for Defense Information, a think-tank in Washington DC. The current US stockpile is far too bloated, he adds.

But supporters of the RRW say that Congress has undercut the weapons programme. Expertise will be lost unless a new generation can design weapons, says Paul Robinson, former director of Sandia National Laboratories in Albuquerque, New Mexico. He notes that the House budget for nuclear weapons is also \$400 million lower than that enacted for 2007. "I feel absolutely sure that members of Congress have not looked at the strategic consequences of what they are doing," he says.

The RRW is not done yet. This week, the Senate begins drafting its own version of the appropriations bill. Senator Pete Domenici (Republican, New Mexico) has vowed to win back support for the programme. A final version of the bill, to be worked out between the House and Senate, is expected sometime in the autumn.

Geoff Brumfiel

HIGH NOTES REALLY ARE HIGH

Perception of pitch and spatial orientation are linked.

www.nature.com/news



Uganda hosts banana trial

A pioneering breed of genetically modified (GM) banana plant could reach a field site in Uganda as early as next month. The tiny plants, currently awaiting transport at the Catholic University of Leuven, Belgium, are expected to become the first GM bananas to enter field trials in Africa. Although unlikely to replace traditional varieties, they may open the door to other GM crops.

The plants have been genetically modified to resist black sigatoka, a leaf spot disease caused by the fungus *Mycosphaerella fijiensis*. The disease can reduce banana production by up to 50%, and so is an attractive target for genetic engineering.

Lab tests show that certain genes from rice make bananas less sensitive to the disease, says Rony Swennen of the Laboratory of Tropical Crop Improvement in Leuven, where the work was done. Whether they perform equally well in the field will be tested at the Kawanda Agricultural Research Institute (KARI) of the National Agricultural Research Organisation, based in Kampala. The fruit, like most cultivated bananas, are sterile, which will keep the genes from spreading to other plants.

After India, Uganda is the world's biggest grower of bananas and plantains, producing 10 million tonnes a year for eating or to make beer. Relatively few of these bananas are of the bogoya dessert variety tested in the Belgian lab — but the same genes may protect them, too.

In fact, the genes may reach Uganda before the bananas: Swennen plans to send the relevant DNA constructs to KARI soon, but the bananas must wait for formal approval from the government. The KARI researchers plan to use the genes to modify highland (cooking) bananas, which are a more important food staple. They also want to improve the highland banana's food value. "We are trying to increase the vitamin A and iron content," says Geoffrey Arinaitwe, a Belgium-trained Ugandan researcher working at KARI.

"The main goal of the project with the bogoya dessert banana is capability building in Uganda," adds Swennen, "so that trained local people can do the work themselves."

The Ugandan project is not the only one aimed at black sigatoka. "We are developing bananas with black sigatoka resistance too," says Eli Khayat, vice-president of research at Rahan Meristem, a biotech firm in Western Galilee,



Genes in bananas might be altered to resist black sigatoka (inset).

Israel. "But we would probably test them in Central America. Treatments against black sigatoka can cost US\$1,600 per hectare in Costa Rica."

The Ugandan project receives help and training from the Agricultural Biotechnology Support Project, an international consortium led by Cornell University. The project is funded by the US Agency for International Development — which also finances a programme advising and supporting the Ugandan government in setting up biotech research projects. "These two big American projects came in with basically one objective: to get GM bananas in the field as fast as possible," says Richard Markham of Rome-based Bioversity International, a non-governmental organization that oversees several banana projects and helped start the Belgian black sigatoka project.

The trials may not draw the public protest that sometimes accompanies GM trials in Europe and elsewhere. "People have more pressing concerns, like the rebellion in the north, AIDS, droughts and poverty," says Markham.

"Some consumer groups strongly oppose GM organisms," adds Godber Tumushabe of the Advocates Coalition for Development and Environment, a Kampala-based think-tank. "But the debate in Uganda has died down in recent years. When field trials begin, it could start again."

Whatever its reception, the banana project still has a long way to go. But it may have smoothed the path for other GM crops. The Uganda government is currently evaluating an application for *Bt* cotton, which produces its own insecticidal toxin, and GM cassava is in the pipeline.

Anne Dauwers

Lost Patagonian lake mystifies geologists

Chile has lost one of its lakes, and nobody knows why.

An unnamed glacial lake, some 20,000 square metres in area, disappeared sometime between March and May in Bernardo O'Higgins National Park in Patagonia, geologists say. One possible explanation is an earthquake that occurred in the neighbouring Aysén region in April, which measured 6.2 on the Richter scale. Such an earthquake might open up a crack in the ground or underlying ice, allowing the water to drain away.

But Stephan Harrison, an expert on Patagonian glaciers at the University of Exeter, Penryn, UK, suspects that an ice dam may have melted, allowing the water behind it to rush out. "This is probably just a natural process," he says.

Geologists are now heading to the site to hunt for more details.

NASA plans how to occupy astronauts on the Moon

In one small step for lunar science, NASA has selected seven proposals for simple experiments that astronauts can carry out on their first trips back to the Moon, currently scheduled for 2020.

The proposals are cheap, costing about US\$100,000 apiece, says Kelly Snook, a lunar project scientist at NASA's headquarters in Washington DC. "They are designed to be autonomous and fairly simple," she says. Nevertheless, Snook adds, they will address some important questions about the Moon.

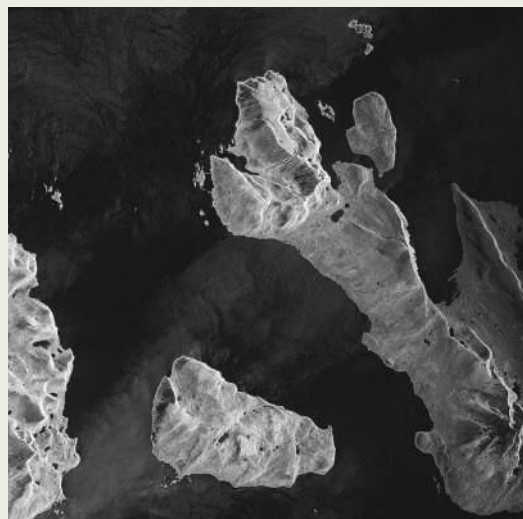
One proposed device, from the Jet Propulsion Laboratory in Pasadena, California, will measure seismicity on different parts of the lunar surface. Others will benefit both science and exploration,

Satellite catches island on camera

Four days after its 15 June launch from the Baikonur Cosmodrome in Kazakhstan, the German radar satellite TerraSAR-X obtained its first detailed images.

Taken for test purposes from 500 kilometres above Earth, this image shows the island of Vanna, 250 kilometres southwest of Norway's North Cape. Its topography and the rough ocean surrounding it can be easily discerned. The bright spots along the neighbouring coastlines are houses.

The satellite, operated under a public-private partnership, will reach its final orbital position by the end of the month. It can visualize objects at up to one metre resolution at any spot on Earth within four and a half days, irrespective of the amount of light and cloud coverage (see *Nature* 446, 22–23; 2007).



GERMAN AEROSPACE AGENCY/LEADS ASTRUM

such as an experiment to look for water and other volatile molecules, from NASA's Goddard Space Flight Center in Greenbelt, Maryland, and a device from Ball Aerospace in Boulder, Colorado, to study lunar dust, which is believed to be hazardous to astronauts.

Stem-cell bill once again falls to Bush veto

For the second time in less than a year, President George W. Bush last week vetoed legislation lifting restrictions on US federal funding for human embryonic stem-cell research, declaring "I will not allow our nation to cross this moral line". At the same time, Bush signed an executive order backing — but not funding — stem-cell research that doesn't require the destruction of embryos, and citing "exciting" work along these lines recently published in *Nature* and *Cell Stem Cell* (see *Nature* 447, 618–619; 2007).

In Congress, supporters of the vetoed bill quickly parried by amending a spending bill in the key Senate spending committee to allow US funding for work on stem-cell lines derived before 15 June 2007. Bush is likely to veto the measure should it reach his desk, leaving further definitive changes in stem-cell policy to the next occupant of the White House.

Research returns to Biosphere dome

The Biosphere 2 terrarium in the Arizona desert once again has a research life, after surrounding land was sold recently for housing.

The University of Arizona at Tucson announced on 26 June that it would lease the 16-hectare Biosphere 2 complex for three years to conduct environmental studies. The Philecology Foundation, set up by Biosphere 2's original philanthropist founder, Edward Bass of Fort Worth, Texas, will fund the lease operations, along with potential gifts and grants from others.

Researchers hope to use the 1.4-hectare dome with an enclosed pond and vegetation for ecological studies. Earlier projects, including the 'biospherians' who sealed themselves inside for a two-year experiment, failed to prove sufficiently productive.

Senate gears up for control over vehicle emissions

As global warming moves to the forefront of the US political agenda, the Senate last week passed a massive energy bill that includes the first rise in vehicle mileage standards in decades.

If it comes into effect, the legislation would require cars and trucks to attain 35 miles per gallon by 2020 — 40% higher than current standards, but a number that would need to be agreed with the House of Representatives later this summer. The Senate also approved programmes for coal-burning power plants that capture and sequester their carbon emissions.

But a proposal to set up a mandatory US registry for companies to log their greenhouse-gas emissions, introduced by Amy Klobuchar (Democrat, Minnesota) and Olympia Snowe (Republican, Maine), failed. And a push to have 15% of US energy come from renewable sources by 2020 also failed.

NASA



What will astronauts do on the next Moon mission?

BUSINESS



The Peterhead project would have converted an old power plant to produce clean energy.

Contest puts brakes on carbon capture

As a BP project to capture carbon dioxide from a power station bites the dust, supporters argue that a major opportunity for Britain has been lost. **Andrea Chipman** reports.

For energy giant BP, it seemed the perfect opportunity to show off its green credentials. Convert natural gas to hydrogen and carbon dioxide at a new plant at Peterhead on the north-east coast of Scotland. Use the waste CO₂ to flush the final reserves of natural gas out of the ageing Miller field under the North Sea, storing the CO₂ in the process, and use the hydrogen to generate clean energy in a second, converted, plant. Show the world that carbon capture and storage can be commercial, and that BP can be part of the solution to global warming.

Yet late last month, the scheme fell apart, after the British government published a long-awaited energy discussion document (white paper) on 23 May. The paper proposed a government competition, starting this November, that would select demonstration projects in carbon capture and storage for later subsidy. The result of the competition was expected towards the end of 2008. BP declared it could not hang around that long and announced that it was dropping the project and closing the Miller field.

Industry watchers say that although the BP project, which was unveiled in 2005, was further advanced than other UK carbon-capture

proposals, its design might have made it more expensive. The project would have cost £500 million (US\$1 billion) to complete, the plant would have had a capacity of 475 megawatts and it would have stored an estimated 1.8 million tonnes of CO₂ annually. BP has already spent around £30 million on the project.

"Peterhead would have been far quicker to bring to fruition," says Sarah Mander, an engineer who works for the Tyndall Centre for Climate Change in England. But "it wouldn't have necessarily" won the competition, she says.

The UK Department of Trade and Industry (DTI) says in the white paper that it wants to see a large carbon-sequestration plant come on stream between 2011 and 2014. But critics say that the competition is not getting under way fast enough to meet this goal.

"If the government is serious about attracting investments from big corporations, it needs to make it attractive to invest," says Stuart Hazeldine, a geologist at the University of Edinburgh and a co-leader of the UK Carbon Capture and Storage Consortium, a

government-funded research group. "BP is going ahead with projects in California and Australia. It is very serious about this."

In the aftermath of the BP decision, Alex Salmond, the newly elected first minister of the Scottish Executive, accused the London-based British government of sabotaging the Peterhead plan. The Scottish Executive formally approved

the Peterhead proposal on 15 June, three weeks after the company had abandoned it, saying that it was keeping the idea alive in the hope that BP might reverse its decision. A company spokesman says it has no plans to do so.

Up to ten other carbon capture and storage projects could enter the competition, says Anthony White, a director of Climate Change Capital, a specialist London investment bank. A 2006 report by the Tyndall Centre and the British Geological Survey estimated the storage potential in the Bunter Sandstone Formation, which covers the southern half of the North Sea, and gas fields to the north at 2.8 billion tonnes and 14.3 billion tonnes, respectively. The report added that a "very large

"A government competition, starting this November, would select demonstration projects in carbon capture and storage."

J.J. MITCHELL/GETTY IMAGES

proportion" of the country's 176 million tonnes of annual CO₂ emissions from power plants could be stored in the Bunter formation alone.

"No other country has a project close to being operational on a commercial scale," says Jeff Chapman, chief executive of the Carbon Capture & Storage Association, a UK trade group. "The UK could get there first."

Small-scale demonstration carbon capture and storage plants are going ahead elsewhere. The German utility company RWE, based in Essen, plans projects in Germany and Britain, and the Swedish power company Vattenfall of Stockholm intends to have a 30-megawatt coal-fired plant operating in Germany by mid-2008. Waste CO₂ is already captured from oil and gas fields in Norway and Algeria. But no one has yet demonstrated the commercial-scale use of the technology at a power station.

Three carbon-capture technologies are on the table. Pre-combustion capture, the model for the Peterhead project, chemically splits methane or gasified coal into hydrogen, which is burned, and CO₂, which is sent underground. It is in use in oil refineries but is more expensive than the other technologies. Post-combustion capture, in which an amine solvent absorbs the CO₂ before it goes up the chimney, has been proved on a small scale and could be retrofitted to existing coal-fired plants. The third option, oxy fuel firing, separates oxygen from the air, releases nitrogen, and burns coal or gas in the pure oxygen: this approach will be used by Vattenfall.

One argument against the Peterhead proposal is that it could not be retrofitted to coal-fired plants, which are responsible for a huge slice of emissions not only in Britain but also in China and India (see page 1038). Those are the nations whose emissions are growing most rapidly, and where Western governments hope that sequestration technology could generate exports.

Whatever the technology, operators are looking for subsidies to meet the high cost of sequestration plant. By the time a prototype carbon capture and storage scheme comes on line, the British government expects to have collected a 'fighting fund' of at least £1 billion for subsidies by auctioning allowances for companies to emit carbon under Europe's emission-trading scheme.

Advocates of carbon capture say the government should concentrate on getting projects going, rather than on selecting which technology to back. Hazeldine claims that Britain has missed opportunities in the past to gain global leadership in technologies as diverse as nuclear power and wind energy. "We've got the legislation and the storage sites and this makes Britain one of the most attractive countries in the world," he says. "The competition is a distraction and a delay." ■

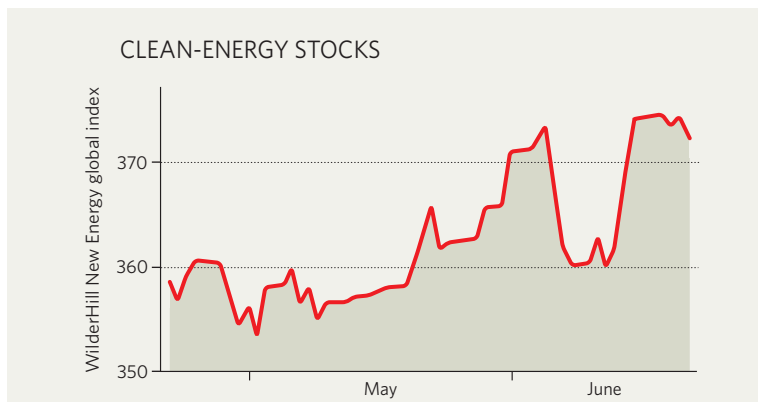
IN BRIEF

TIGHTENING UP A congressional committee has unanimously approved legislation that would give the US Food and Drug Administration (FDA) significant new powers to police the safety of drugs that are already approved and on the market. The legislation passed by the House Committee on Energy and Commerce on 21 June is similar to a bill already approved last month in the Senate (see *Nature* 447, 247; 2007). Neither contains controversial provisions creating a path for speedy FDA approval of generic copies of biological drugs — leaving these to be considered separately, probably later in the year.

PATENT PERMITTED The CSIRO, Australia's main research agency, has won a landmark court case in Texas that appears to restore the rights of such agencies to claim infringement of patents. A judge in Tyler upheld the research agency's complaint against Buffalo, a Japanese telecommunications company, regarding wi-fi technology patented by the CSIRO. The case is seen as significant for other patent-holders because it dispels concerns that, after an unrelated US Supreme Court ruling last year, infringement can only be claimed by parties that are actually selling products based on their patents.

TAKEOVER TALK Shares in Bristol-Myers Squibb rose by 4% after a New York court ruled that the patents on its blood-thinning drug Plavix were good for another four years. The ruling on the world's third best-selling drug fuelled speculation that the company could be the target of a takeover bid (see *Nature* 446, 15; 2007). The patents on Plavix are held by French drug company Sanofi-Aventis, but Bristol-Myers Squibb has exclusive rights to sell the drug in the lucrative US market.

MARKET WATCH



Stocks in renewable-energy companies around the world have continued to rise this spring, confounding widespread expectations of a market correction — at least for the time being.

The WilderHill New Energy Global Innovation index (symbol NEX on the American stock exchange) hit record highs again this month (see graph) and the basket of stocks it contains is now worth almost four times as much as it was in 2002. With investors reconciled to continuing high oil prices — perhaps indefinitely — and government and corporate initiatives on 'clean energy' cropping up every other day around the world, investors in the sector have never had it so good.

According to Rob Wilder, president of WilderShares, a California consultancy that co-compiles the index with New Energy Finance of London, the latest

boost arrived in the shape of a fresh energy bill being drafted in the US Congress. Although its final shape remains to be seen, it is likely, Wilder says, that the measure will divert "billions of dollars" in subsidies from the oil industry to renewables.

There has been some takeover activity. On 20 June, electronics giant Philips purchased Color Kinetics in Boston, which makes lighting systems based on light-emitting diodes (see *Nature* 447, 766–767; 2007), for US\$790 million. But an unusual characteristic of the boom is the relative lack of consolidation among clean-energy companies. "There's almost the opposite, with new companies sprouting up all the time," Wilder says. "Every company thinks they have a shot to make it. If I ran a company in this sector, I wouldn't want to sell it." ■

Colin Macilwain

ENERGY-GO-ROUND

How did a little Spanish province become one of the world's wind-energy giants? **Daemon Fairless** reports.



From a wheatfield overlooking the village of Iratxeta in the Spanish region of Navarre, you can appreciate the unique countryside and its contrasts: the pale green of ripening grain setting off the dark-green mountain forests behind, country houses newly rebuilt on medieval foundations, and giant wind turbines, brilliantly white and strikingly erect, their slowly moving blades driven by the breeze that ripples the wheat.

"That was the first one," says José Roman Gómez, pointing south to the Guerinda wind park, a phalanx of turbines running along the ridge of a nearby hill. He draws an arc along the horizon with his finger: "That one was built after. And there is a different one. And, over there, another." Iratxeta, an isolated cluster of stone houses nestled into the hills about 30 kilometres south of the regional capital, Pamplona, is encircled by wind farms. Their turbines loom over the hills like an invading army. But it's an occupation for which Gómez is grateful.

Gómez, who has been managing the local council of the villages of Iratxeta and neighbouring Leoz for the past two decades, is a tranquil and friendly man in his mid-50s. Fifteen years ago, before the wind boom hit this area, the two villages had only about 150 residents between them. "There was no running water here," Gómez explains. "There was no waste collection or buses. Most of the villagers were old and waiting to die."

Now, however, the villages have nearly doubled in size compared with before the wind boom. In addition to providing running water and waste collection, the rent from wind-farm operators has enabled the local council to hire four full-time health workers to care for the community's elderly people. "Now," Gómez says, "people from Pamplona are investing here."

Wind farms have transformed Spain's landscape over the past decade and Navarre, the least populated of the nation's 15 mainland autonomous regions, has been at the forefront. Located in the northeast, between the Basque country and the French border, Navarre generates almost 60% of its electricity from renewable sources — the vast majority from wind. The region's most recent energy plan, released in May, aims to increase this proportion to more than 75% by 2010.

From water to wind

Although it would hardly be fair to attribute Navarre's renewable-energy boom to just one person, one name invariably arises when discussing wind power, that of Estaban Morrás.

Morrás is the executive director of Acciona Energía, a subsidiary of the Spanish engineering group Acciona, which has grown to be the world's foremost wind-park developer, accord-

ing to the Madrid-based Spanish Wind Energy Association, an industry body. Dark-suited and seated behind a boardroom table at the company's headquarters on the outskirts of Pamplona, he looks very much the person you would expect to see running a company with first-quarter revenues in 2007 of €252 million (US\$337 million). But when Morrás founded the company in 1989, he had more of an Indiana Jones air about him. "I did so much research, running through rivers and forests looking for locations for new hydroelectric projects, my sweaters, my clothes, disintegrated," he says.

Until the early 1990s, small, decentralized hydroelectric stations dotting the rivers that flow out of the Pyrenees were the only significant form of renewable energy in Navarre. Many had been around since early last century, providing power to local small industry.

Morrás founded his company, originally called EHN, with the plan of buying up the existing mini-hydro-stations, connecting them to the grid and building more. But it was becoming harder to find new sites. "I spent all year in 1992 desperately looking for new locations," he says. "But it was impossible. I understood my company had no future." But, in 1994, on a trip to France, he saw the future.

"The most significant moment for me," he recalls, "was visiting a wind farm in Montpellier. My first thought was that a single turbine was capable of producing more energy than the last mini-hydro plant we had built. It was really difficult for me to sleep that night." He took to the new technology with gusto and ambition. Morrás says that in his first meeting in the mid-1990s with Vestas, the Danish company that he selected as his turbine supplier, he told the then chief executive, Johannes Poulsen, that Spain would have 1,000 megawatts of wind-power capacity by the end of the century. At the time, it had virtually none.

Poulsen, he says, was incredulous: "You are crazy," he told me: "Denmark has only 400 megawatts after 20 years of effort." Morrás's answer was "with Denmark's experience, we can go faster than you." And they did. With 11,615 megawatts of installed wind capacity at the end of last year, Spain is second in the world to Germany in capacity, beating the United States by just 39 megawatts (see graphic, overleaf).

The energy actually generated is much less, however, thanks to the caprice of the wind and maintenance downtime for the turbines. The

23,372,000 megawatt-hours generated in 2006 is only 23% of the turbines' full capacity. Nevertheless, it provided just over 8.5% of Spain's electricity: only Denmark boasts a higher percentage. On a particularly blustery day this March, wind generation was providing 27% of Spain's electricity supply — surpassing, for a fleeting moment, the contributions of both nuclear and coal-fired plants.

Power to the people

The wind boom in Navarre is among the most intensive in Spain. Throughout the 1980s and 1990s, the government of Navarre was eager to develop new industries, especially as the region seemed overly dependent on its single large industrial employer, a Volkswagen car plant. But with a limited local power supply, Navarre's infrastructure wasn't particularly appealing to outside investors.

Morrás's wind-power plan offered a solution, and promised an industry in its own right — more enticing for the government at the time than any environmental argument.

Well before Spain's national government started to guarantee profitable prices for selling wind-generated power to utilities, Navarre had bought into Morrás's vision. His first wind

farm, El Perdón, was built south of Pamplona in 1994; since then the government of Navarre has approved well over a thousand turbines in 32 wind farms. From 1995 to 2004, it invested more than €136 million in renewable-energy enterprises, contributing up to 30% of the initial funding and providing tax credits for investors. The current renewable-energy plan (2005–2010) has earmarked an extra €240 million for investment in renewables. Navarre's installed wind capacity is about 950 megawatts — 8.5% of the country's total on 2% of its land, and nearly two-thirds that of its neighbour, France.

For a while it looked as though Navarre might become entirely self-sufficient by 2010, producing as much electricity through renewables as it consumed. "Our goal in this area was more aggressive than it is now," admits José Javier Armendáriz, the region's minister of industry, technology, commerce and labour.

Morrás's aspirations haven't changed, however. "I think it's possible for Navarre to obtain 100% of its energy with renewables," he says. "This project is essentially one of companies and people in the private sector. I don't understand the position of the government. The target



for me is to obtain 100%.” His goals to expand renewable energy go well beyond the region. “I am convinced that it is possible for the world to have an energy model that is 100% renewable.”

Navarre could do nicely out of this. By 2003 the region had more than 50 companies active in various aspects of renewable energy (J. Faulin *et al. Energy Policy* **34**, 2201–2216; 2005), and that year it was commended as having the best regional policy in Europe at the European Conference for Renewable Energy in Berlin. EHN became lucrative enough to attract the attention of the Acciona Group, which after first taking a small stake, took it over entirely in 2005. The Navarrese wind boom also gave birth to Gamesa Eólica in Pamplona, initially created as a local supplier for EHN, which has grown to be one of three turbine producers vying with each other for second place to Vestas in the world turbine market.

Pamplona is on its way to becoming the official hub of Spanish renewable energy. The city is home to the National Renewable Energy Centre, a new research facility that focuses on developing commercial applications in wind, solar and bioenergy, and to two national training facilities in renewable technologies. Last autumn, the Public University of Navarre in Pamplona launched the country's first graduate programme for electrical engineers in wind and solar electricity, to meet the growing demand for specialists.

Running out of land

The region's success has led it into uncharted waters, however. Although much of the rest of the world is still waking up to renewables, Navarre has reached possible saturation point, at least for wind power. In 2004, the regional government stopped approving proposals for new



Solar panels and wind farms have helped to regenerate the entire Navarre region.

wind farms. One reason, it says, is the effect on the landscape. According to Armendáriz, the government sensed that further development might lose it the public support that, he says, was one of the reasons wind power could gain such traction in the region.

Another reason is that wind-power capacity can grow without building new farms, by installing newer, higher-capacity turbines. Several of the region's first wind parks, those with 500-kilowatt machines, will be fitted with newer 3-megawatt models, starting this summer.

Even so, having so much wind energy contributing to the network comes with technical concerns. The amount of energy that each wind turbine produces depends entirely on what the wind is doing, and can vary from moment to moment. This makes it difficult for power utilities, which need to match the amount of energy generated with the amount being consumed moment-by-moment.

If you rely on wind power, you need spare capacity for when the wind drops, and in 2004, when new farms stopped being developed, the regional government approved two 400-megawatt natural-gas plants near Castejón, a town in southern Navarre. Luis María López González, an engineer specializing in renewable energy at

the University of La Rioja in Logroño, suggests that the intermittent nature of wind generation was one of the reasons that the plants were built. These ‘combined-cycle’ plants, which harness their own waste heat, can be fired up quickly to meet a sudden need. They are also efficient, emitting a little over half of the carbon dioxide per watt that a coal-powered plant does. But that's still a lot more carbon dioxide than a windmill.

With the wind sector maturing, Spain as a whole is starting to look elsewhere for renewable-energy sources. Last month, the national parliament lowered the subsidies that wind operators get and boosted subsidies for the production of biomass and biofuels and solar photovoltaics. One new scheme is so-called ‘solar gardens’, large expanses of solar panels in which individual investors can buy as few as a single panel, enjoying both the income it provides and the feeling that they are greening the grid.

Morrás has no complaints about the government shifting funds away from wind and into other forms of renewable energy, even though it means that making a profit from wind is harder. “It will be an effort for the sector,” he says, “but we can do it.” And there is still plenty of room for wind-farm growth elsewhere. “In my opinion,” he says, “the next years will be the most important, with the extension of wind farms in China, India, South America, Europe and Russia.”

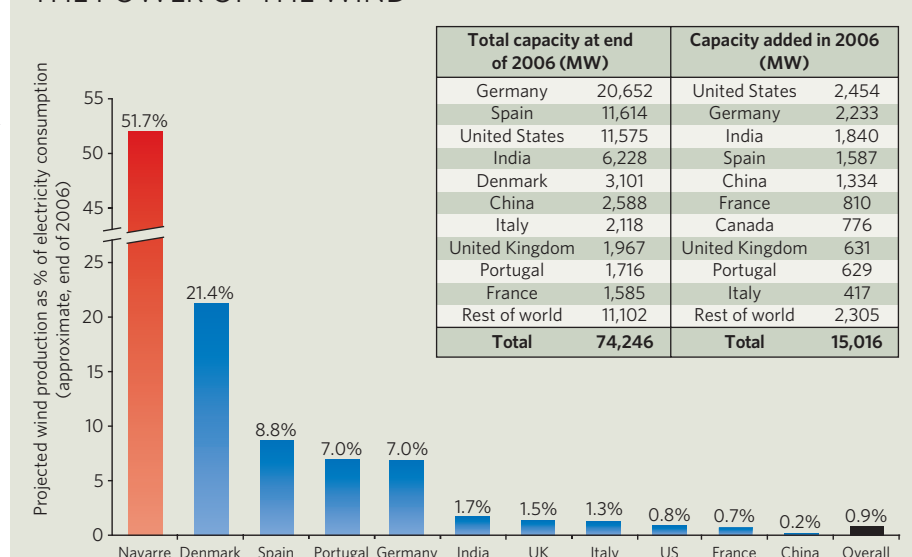
Closer to home, in Iratxeta, Gómez has been overseeing a more modest expansion — the erection of a new building for the village council, one of several infrastructure projects the village has invested in with the income they receive from the wind parks.

For Gómez, wind power is a renewal as well as a renewable: it has brought Iratxeta and its neighbouring villages back from what seemed to be an inevitable decline. He is disappointed with the government's decision to make no new wind parks. “I think if you ask people around here,” he says, “they'll tell you they want more because it means more money. Wind power,” he says with the emphasis of someone who has seen science change his world for the better, “has been a copernican revolution.”

Daemon Fairless is this year's winner of the IDRC-Nature fellowship.

D. FAIRLESS

THE POWER OF THE WIND





Out from the haze

Smoking was banned in Californian bars a decade ago, and this week England follows suit. But **Kris Novak** finds that epidemiologists are still arguing about the effects of second-hand smoke.

The reactions to the English ban on smoking in all enclosed public places on 1 July are almost as predictable as a chain smoker's second pack of the day. "First you'll read that all the pubs are going broke, then that people are blocking the street when they stand outside smoking, creating traffic congestion, and then everyone will be told that the ban is wildly unpopular," says Stanton Glantz, director of the Center for Tobacco Control Research and Education at the University of California, San Francisco. "It happens every time these bans go into effect. After the wave of bad press, everything will calm down and next year everyone will say, 'What was the big deal?'"

Glantz should know — he was involved in the first widespread ban on smoking in workplaces, including bars and restaurants, which has been in effect in California for almost ten years. Lately, he has advised public-health officials from around the world on how to enact similar legislation.

Back in 1998, banning smoking in bars was not easy, even in clean-living California. The ban met with resistance from groups claiming that making bars smoke-free would devastate businesses, deny adults the freedom to smoke and be too difficult to enforce. The tobacco

industry made nine attempts to repeal the law and spent US\$18 million on a public-relations campaign against it.

However, a smaller but well-organized public-health campaign overcame resistance to the ban. Glantz says that it's important that governments explain the dangers of second-hand smoke to non-smokers. In his view, passive smoking has been the Achilles' heel of tobacco companies, taking the focus off the rights of smokers and placing it on the health of non-smokers. True to this script, over the past six months, the UK government's health department has spent £8.7 million (US\$17.3 million) on an advertising campaign to educate English citizens about second-hand smoke and prepare businesses for the change.

Coronary concerns

Despite all the strongly worded advertising campaigns, the health benefits of smoking bans for non-smokers have been controversial. One of the biggest rows concerns the links between second-hand smoke and heart disease — an argument in which Glantz has been a key player. Richard Peto at the University of Oxford, who has been studying tobacco use for more than 30 years, says: "Passive smoking must kill some people, but the big question is how many."

Smoking has been banned in all workplaces, including bars and restaurants, in more than 20 countries (see table) and in many US states. In March 2004, Ireland became the first country to enact such a ban. "Ireland is a huge success story," says Glantz, paving the way for bans in Scotland, Wales, Northern Ireland and now England, the largest nation to do so. Glantz cited the active support of Ireland's health minister and television campaigns that warned people of the dangers of second-hand smoke.

Because California has the longest history of smoking restrictions, epidemiologists often turn to the state's public-health data to determine the effects of smoking bans, on both smokers and non-smokers. Back in 1988, California voters approved the California Tobacco Control Program (CTCP), which increased tax on cigarettes to fund tobacco education programmes, and went on to ban smoking in workplaces (including restaurants) in 1995 and in bars in 1998.

Several studies show that the CTCP helped Californian smokers to quit. In 1988, 23% of all Californians smoked and by 2006 that number had dropped to 13%¹. Similar effects have been seen in Europe: in Scotland, cigarette sales fell by 8% in the first year of its ban, and Ireland has seen about a 5% drop in the number of

S. DIONISIO/AP/PA

smokers since the 2004 ban, although 25% of Irish people still smoke.

But were these changes on their way, with or without a ban? Peto notes that smoking-related deaths in the United Kingdom have already declined by almost 50% since the late 1960s. "Smoking bans will probably affect some additional people," he says, but separating that effect from the overall decline in smoking is a complex task.

The California experience suggests that legislation has a role. More than half of California's former smokers agreed that the smoke-free workplace and bar laws made it easier for them to quit, and most smokers reduced their cigarette consumption because of the bans¹. According to the European Commission, the United Kingdom has the highest percentage of European smokers who are trying to quit (46%), so the 1 July ban could be a step in the right direction.

Tips for non-smokers

If smokers are giving up in droves, is their health also improving after the bans? Once again, the California experience is helpful because public-health data can be compared with those for US states without smoking bans. Between 1988 and 2002, rates of lung and bronchial cancer declined four times faster in California than in the rest of the United States¹.

Peto, however, is sceptical of studies that associate mortality trends in large populations with a single event. "The number of deaths from vascular disease and cancer have been falling for multiple reasons, such as improved treatments; at the same time, you have increases in disease rates because of obesity — there are just too many things going on to attribute changes in mortality to one single thing," says Peto. He thinks that it takes decades for population-wide changes in health trends to become apparent.

But what about the health benefits of smoking bans for non-smokers? These are even more uncertain. Banning smoking does clean up the air in bars — an Italian study comparing air quality before and after their 2005 ban found a 95% reduction in airborne nicotine². Within months of bans on smoking in bars in Ire-

Country	Year of smoking ban
Argentina	Regional bans in 2006 and 2007
Australia	Regional bans in 2006 and 2007
Bhutan	February 2005
Denmark	April 2007
England	July 2007
Finland	June 2007
France	Scheduled for January 2008
Hong Kong	January 2007
Hungary	Scheduled for 2009
Iceland	June 2007
Ireland	March 2004
Italy	January 2005
Lithuania	January 2007
Netherlands	Scheduled for July 2008
New Zealand	December 2004
Northern Ireland	April 2007
Norway	June 2004
Portugal	May 2007
Scotland	March 2006
Sweden	June 2005
Thailand	August 2007
Wales	April 2007
Uruguay	March 2006

land and Scotland, bar workers reported fewer respiratory problems, and breath, saliva and blood samples contained fewer biomarkers of tobacco exposure than before the bans^{3,4}.

According to Smokefree England, an educational website funded by the UK health department, second-hand smoke contains more than 4,000 different chemicals, such as carbon monoxide, and more than 50 carcinogens, including chromium, vinyl chloride and benzene. So does cleaner air prevent disease in non-smokers?

Smokefree England claims that exposure to second-hand smoke increases a non-smoker's risk of getting lung cancer by 24% and risk of heart

disease by 25%. But where do these numbers come from? "These are generally accepted numbers," says epidemiologist Alfredo Morabia, at Queens College in New York, who cites the 2006 US surgeon general's report, *The Health Consequences of Involuntary Exposure to Tobacco Smoke*, the US government's most detailed statement ever on second-hand smoke⁵.

For lung cancer, there is general agreement with the report's assessment. Studies of non-smokers who live with a smoker suggest that the risk of developing lung cancer is about 25 times higher in smokers than non-smokers. Because non-smokers' exposure to second-hand smoke is estimated to be about 1% that of a smoker, a 24% increase in lung-cancer risk (0.24-fold) fits with expectations.

Cardiac effects

But the data supporting the link between second-hand smoke and cardiovascular disease are more controversial. The surgeon general's report states that "pooled relative risks from meta-analysis indicate a 25–30% increase in risk of coronary heart disease from exposure to second-hand smoke". Although most epidemiologists think there is a link, it's the size of the effect that surprises them.

"It seems to me that a 25% increase is not plausible," says John Bailar, a biostatistician at the National Academy of Sciences in Washington DC, who thinks the effect should be proportional to exposure, as it is for lung cancer. "Regular smoking only increases the risk of cardiovascular disease by 75%, so how could second-hand smoke, which is much more dilute, have an effect one-third that size?" Bailar says that even if a non-smoker took in 10% as much smoke as a smoker, which is a high-end estimate, his increased risk would be only 7.5%.

One explanation offered by Glantz and other heart researchers for the higher-than-expected effect of second-hand smoke on coronary heart disease is that the 'sidestream' smoke a non-smoker breathes is more toxic, per gram of total particulate matter, than the 'mainstream cigarette smoke' that a smoker inhales⁶. So small amounts of sidestream smoke might be more likely to trigger heart disease than the smoke inhaled through a filtered cigarette.

Another explanation is that a very low threshold of exposure to second-hand smoke is required for disease risk to escalate. But this idea is hard to test. Epidemiologists lament that exposure is one of the hardest factors to quantify. "There is no unit of exposure, and levels of exposure vary based on what size room people are in, the number of smokers in the room, the



Campaigns against passive smoking often cite data from the 2006 US surgeon general's report (top).

level of ventilation and so on," says Morabia.

Despite these concerns, the surgeon general's report takes a hard line on exposure, stating that there is no 'safe' level. According to Terry Pechacek, one of the authors of the report and associate director at the Office on Smoking and Health at the US Centers for Disease Control and Prevention in Atlanta, Georgia: "Exposure to second-hand smoke for even a short time can have adverse health effects — this is not subject to debate. Compounds in tobacco smoke have the ability to cause cancer in humans, it's just a probabilistic game of whether they will cause death in a certain individual."

Burning opportunity

In principle, smoking bans provide a unique opportunity to study populations before and after reductions in second-hand smoke exposure. But if research into the incidence of heart attacks is any guide, the results of such studies are often far from clear and can cause more controversy.

Glantz reported in 2004 that during a six-month smoking ban in Helena, Montana, the number of heart attacks dropped by 40% compared with the same months in other years⁷. The study, cited in the surgeon general's report, was criticized for the small number of cases studied and the large month-to-month variations in incidence of heart attacks.

Since the report, several studies have reported a drop in hospital admissions for heart attacks after smoking bans in the Piedmont region of Italy (11% drop)⁸, and in small cities in Colorado (27%)⁹ and Ohio (39%)¹⁰. But not all epidemiologists are impressed. "It's quite common to see major year-to-year changes in heart attacks, sometimes as much as a 50% increase or decrease," says Michael Siegel, an epidemiologist at Boston University School of Public Health. Attributing such changes to a ban is impossible, he says.

Worse, none of the studies recorded whether the changes occurred in non-smokers or in smokers, yet the effects of the ban are frequently attributed to reductions in second-hand smoke exposure. The lead author of the Ohio study agreed that a prospective study that collected

data on groups of non-smokers and smokers in advance of a ban — and then followed them up for a few years after — would be ideal, but this would take longer and be costly.

Glantz has heard all these concerns before: "While the numbers were small in our original Helena study," he says, "the more recent ones have been in larger places." In his view, the effects have now been well documented by several studies, all of which accounted for monthly variations in heart attacks: "That is not a legitimate criticism of any of them."

Lethal limits

Certainly, public-health officials seem convinced. Pechacek says that the general consensus in tobacco researchers is that the drop in heart attacks occurs mainly in non-smokers. Smokefree England suggests that just 30 minutes of breathing in second-hand smoke can raise your risk of having a heart attack. But can you really become ill just by sitting in a bar next to a smoker? "Saying that just a little exposure is killing people is going overboard," says Siegel, who worries

that when researchers exaggerate their findings, they lose credibility with the public. "I agree that second-hand smoke is a tremendous health hazard, but no one is going to have a heart attack from 30 minutes of exposure."

Siegel thinks that banning smoking in outdoor places is going too far and risks losing support for smoking bans overall. Smoking has been banned on 25 California beaches, and this week the Beverly Hills City Council approved a ban in nearly all outdoor dining areas. "We should focus efforts on the remaining areas in which workers are not protected," he says. "My biggest concern is for the waiters and bartenders who spend 40 hours each week in very smoky environments." He's also worried about the health of smoking research itself, which he sees being compromised by methodological flaws and over-interpretation of results.

And although prospective studies would be costly, Siegel argues that funding for a large prospective study of the effects of smoking bans on non-smokers is warranted. "Although I wholeheartedly support smoking bans, I still believe that we must use solid science to advocate

for such bans and that a noble end — improving public health — does not justify the compromise of our scientific principles."

Other epidemiologists worry that with all the focus on second-hand smoke, the attention is being drawn away from the real issue: the dangers of smoking itself. "The key point to remember is that smokers kill themselves," says Peto. "A few are probably killing other people, but half of all smokers will be killed by their own tobacco."

Kris Novak is a freelance writer based in California.

1. California Tobacco Control Update (California Department of Health Services Tobacco Control Section, 2006).
2. Gasparri, A. *et al. Epidemiol. Prev.* **30**, 348–351 (2006).
3. Menzies, D. *et al. J. Am. Med. Assoc.* **296**, 1742–1748 (2006).
4. Goodman, P. *et al. Am. J. Respir. Crit. Care Med.* **175**, 840–845 (2007).
5. *The Health Consequences of Involuntary Exposure to Tobacco Smoke — A Report of the Surgeon General* (US Department of Health and Human Services, 2006).
6. Schick, S. & Glantz, S. *Tobacco Control* **14**, 396–404 (2005).
7. Sargent, R. P., Shepard, R. M. & Glantz, S. A. *Br. Med. J.* **328**, 977–980 (2004).
8. Barone-Adesi, F. *et al. Eur. Heart J.* **27**, 2468–2472 (2006).
9. Bartecchi, C. *et al. Circulation* **114**, 1490–1496 (2006).
10. Khuder, S. A. *et al. Prev. Med. Advance online publication* doi:10.1016/j.ypmed.2007.03.011 (2007).



Bye to The Smoke? Scientists say a ban on smoking outdoors is going too far.

**"We must use solid science to advocate for such bans."
— Michael Siegel**

Researchers fight poaching with presence, not guns

SIR — Your News Feature ‘Peaceful primates, violent acts’ (*Nature* **447**, 635–636; 2007) reports on the conflicts that arise when wild animals studied for research are threatened by poaching and the bush-meat trade. Regional and international conservation organizations can help, but sometimes individual researchers feel that more immediate measures are required. Local presence has been shown to be one of the most efficient conservation actions, and many research programmes, including the bonobo research project of the Max Planck Institute, have taken risks in continuing to work even when unrest prevails.

You tell the story of Jonas Eriksson, a PhD student who left his academic career to engage in an unusual form of conservation action. From your report, readers may have gained the impression that Eriksson has been engaging in firefights using guns obtained illegally, but this was not the case. The aim of the project was to strengthen the capacity of the guards of the Congolese wildlife authority (ICCN) and to lead joint patrols of villagers and park guards into areas of Salonga National Park where poachers operate. The guards from ICCN are armed with automatic weapons that are owned by the wildlife authority, with a mandate to use them for law enforcement.

We emphasize that the anti-poaching project is neither typical nor representative of the work of the Max Planck Society. Researchers at LuiKotal, in Salonga National Park, have never been armed. Carrying arms would violate national and international laws, and would be counterproductive to the goals of our research.

Conservation and research have to go hand in hand, without weapons. The pressures that we can exert are physical presence and a strong motivation to protect those who provide us with the information we seek. This is what Eriksson did when he started his Salonga mission, and it is what other researchers from our institute do at their field sites across the African continent.

Gottfried Hohmann

Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany

Estimating the cost of climate change

SIR — Your News in Brief story ‘Germany counts the costs of climate change’ (*Nature* **446**, 360; 2007) reports the conclusions of a study by the German Institute for Economic Research (DIW) that you state is based on my

work. This is not correct. As I had previously documented (R. Roson and R. Tol *Integr. Assess.* **6**, 75–82; 2006), the DIW model is based on egregious misinterpretation of my work.

The German gross domestic product (GDP) was about €2,000 billion (US\$2,700 billion) in 2006. If the German economy grows by 1.65% per year without climate change and 0.5% slower with climate change, as the DIW study suggests, then the gap between the two scenarios is the €800 billion in 2050 that you reported. This corresponds to 20% of German GDP in 2050, without climate change. This is at the upper end of the range of the *Stern Review*. However, you call the DIW study “less pessimistic” than the *Stern Review*.

You are correct to say that my estimates of the impacts of climate change are lower than those of the *Stern Review*, and, by implication, the DIW study. Indeed, as I showed earlier (R. Tol *Energy Policy* **33**, 2064–2074; 2005), peer-reviewed estimates are lower than estimates in the grey literature. Neither the DIW study nor the *Stern Review* were reviewed by independent peers.

Richard Tol

Economic and Social Research Institute, Whitaker Square, Sir John Rogerson’s Quay, Dublin 2, Ireland

Not so sunny view of the events in Arizona

SIR — You seem so bedazzled by President Michael Crow’s brutal reshaping (‘The Arizona Experiment’ *Nature* **446**, 968–970; 2007) of Arizona State University (ASU) that I could not help thinking of Stephen Poliakoff’s *Blinded by the Sun*, a play depicting the cold-fusion debacle. Your comment “Take Robert Pettit — a chemist and long-time director of the Cancer Research Institute at ASU until he lost the position in 2005” hardly touches on the ferociousness of President Crow and the events you describe. John C. Knight, in Correspondence, described those events differently (*Nature* **447**, 528; 2007).

Pettit is probably one of the most prolific and productive scientists at ASU. He has published more than 750 articles in the cancer field, and was responsible through his own prodigious funding efforts for the construction, maintenance, operation and staffing of the institute. The efforts of Pettit, his students and collaborators are known worldwide, thanks to their investigations during the past 30 years of the anti-cancer properties of natural marine products. The institute brought more funds to ASU, through its patent income, than any other inventor.

Pettit’s institute was closed with startling lack of notice and the staff of more than 60

people were marched out of the building on 27 January 2006 by security personnel (see *Chemical and Engineering News* **10**, 6 February 2006). In terms of harsh abruptness, this step must be unprecedented in US academic history.

A pending, multi-million dollar suit against Crow and ASU filed by the Government Accountability Project with the US District Court in Phoenix may yet shed light on this sorry affair.

More often than not, excessive sunshine produces sunburn, some of it even fatal.

Carl Djerassi

Department of Chemistry, Stanford University, Stanford, California 94305-5080, USA

Why are people reluctant to join in open review?

SIR — I was excited when *Nature* launched its trial for open peer review last year, but disappointed by the outcome (see <http://blogs.nature.com/nature/peerreview/trial>). I have also been tracking the progress of another open review journal, *Biology Direct* (www.biology-direct.com; see article in *Nature*’s peer-review debate at www.nature.com/nature/peerreview/debate/nature05005.html). Even after a high-profile launch with guaranteed indexing by PubMed, this journal has published only 52 articles and received only two comments over 16 months.

In contrast to *Biology Direct*, another journal that offers open commenting, *PLoS One* (www.plosone.org), has published 1,189 articles in its first six months. But has *PLoS One* achieved its stated goal of post-publication open comments? I find that even the ‘most annotated’ category of articles usually receives just a few comments. The journal has recently replaced its ‘most annotated’ with a ‘recently annotated’ category. A check of all ‘recently annotated’ articles demonstrates that their commenting rates are low (zero or just a few), even for articles that are likely to have broad appeal and/or are in ‘hot’ research areas.

Why is there a general lack of interest among the scientific community in open commenting on submitted or published papers? I believe there are two main reasons. First, participation does not earn any tangible credit or benefit for the reviewers and commentators. Second, publicly critical comments are a risk for those who make them.

Shi V. Liu

Scientific Ethics, Apex, North Carolina 27502, USA
<http://im1.biz>
SVL8SE@im1.biz

Science publishing issues are regularly featured at Nautilus (<http://blogs.nature.com/nautilus>), where debate is welcomed.

COMMENTARY

Tony Blair's era: good news, bad news

As Gordon Brown becomes Britain's prime minister, Robert May highlights some critical challenges in the continuing support of science.

The Blair decade has in many ways been a good one for UK researchers in science and engineering. From the outset of his premiership, Tony Blair has emphasized the importance of scientific research for understanding the natural world and our place in it, and for the practical benefits it brings. And he meant it: during the early years of his office the rate of increase of the Office of Science and Technology's budget was among the highest of all government departments. Such funding patterns have helped to diminish, though not close, the gap with other leading countries. They derive from the belief of Blair and his chief finance minister and imminent successor Gordon Brown that, in Blair's words, "the science base is the absolute bedrock of our economic performance"¹.

The government has also addressed problems in the transfer of knowledge from academia to the private sector, through various programmes. The idea has been to encourage the kind of productive inventiveness that characterizes places such as Silicon Valley in the United States. Signs of success here over the past 10 years can be read in the greater number of companies spun out from universities, ownership of patents and much else.

This is all good news. So why is there discontent among scientists in UK universities? Are we perhaps revealing an inherent, Eeyore-like glumness? I think not.

Although student numbers, faculty positions and research funds have all increased, they have not done so in equal proportions. Student-to-staff ratios are almost universally higher than 15 years ago. And research funding — despite its increase in real terms — has not kept pace with the increasing number of active researchers, themselves spurred on by the demands of the Research Assessment Exercise (RAE), which every few years rates individual academic departments. Life seems harder than it used to be.

Even more important, in my opinion, than higher teaching loads and the increasingly fierce competition for grants, is the extreme growth of bureaucracy — too often masquerading as accountability. The ballooning of the civil service since 1997 means that there are now more conscientious administrators who hold meetings and send out forms to be filled in. And universities have matched or exceeded



Will Gordon Brown have any bright ideas for UK science?

the growth in bureaucrats seen in the civil service. This growth is only partly justified by need. One issue that Brown might address is that the current number of central administrative staff is roughly equal to the number of faculty for four major UK universities (J. Enderby, personal communication); this would certainly raise eyebrows at many top US universities.

The pressure's on

A rather different issue that has emerged during the Blair decade is the tendency to invite people from the world of business to advise on the management of universities, or to head them. Given that UK universities still stand significantly higher on international league tables than does most of the UK business sector, this seems odd. In a recent study², Amanda Goodall of Warwick University looked at 55 UK research universities and showed that improvement in RAE scores (from 1992, 1996 and 2001) is significantly correlated with their head's academic record. I think there is an important lesson to be learnt here.

Above all, in the interests of science and for the greatest returns on our investment in it, I would urge Brown's government to focus on changing the RAE. I believe that it is creating pressures for every department in every university either to be doing front-rank

research or to be closed down.

A few statistics show how inappropriate this is. Scaled against population size, for every 100 PhDs produced in the United States in 2001 in science, medicine and engineering, the United Kingdom produced 155 (in both cases, just over one-third were obtained by foreign students). This surprising figure is explained by Britain's large number of PhD-granting institutions: the United Kingdom, with one-fifth the population and one-sixth the gross domestic product of the United States, has as many as 120 such institutions, compared with around 270 in the United States.

I believe this shows that there are too many pressures to have PhD programmes at the cost of providing higher education for a diversity of ambitions and applications. I am not for a moment advocating anything like the pernicious division between research universities and teaching universities proposed some years ago, but rather for policies that promote diversity of aspiration within the higher-education sector.

Excellence in research is dispersed widely and unevenly within and among universities in Britain. One study³, which analyses citations per faculty member, identifies the top three UK departments in each of 21 areas of science, medicine and engineering. In doing so, it recognizes 63 departments, distributed among 27 universities (with the top one achieving 11). More explicit totems — and rewards — in addition to the RAE are needed, to recognize excellent teaching, the employability of graduates, usefulness to local communities and other achievements.

Reflecting, at this point of transition in Britain, on a happy and lucky life that I have spent as a researcher and teacher in universities in Australia, the United States and the United Kingdom, I think Britain always has been, and still is, an excellent place to be. But I worry that many current trends are potentially damaging. Unless Gordon Brown addresses these issues, I believe the country is less likely to thrive scientifically over the next decade. ■

Robert May was chief scientific adviser to the UK government 1995–2000 and president of the Royal Society 2000–2005.

1. Blair, T. *Science* **281**, 1141 (1998).
2. Goodall, A. J. *Doc.* **62**, 388–411 (2006).
3. Anon. *Science Watch* **8**, 1–2 (Jan/Feb 1997).

K. WIGGLESWORTH/AP/PA

BOOKS & ARTS

Falling over the edge

Claims that an intelligent designer is needed to explain evolution of complex systems are deeply flawed.

The Edge of Evolution: The Search for the Limits of Darwinism

by Michael Behe

Free Press: 2007. 336 pp. \$28

Kenneth R. Miller

Michael Behe's new book, *The Edge of Evolution*, is an attempt to give the intelligent-design movement a bit of badly needed scientific support. After a spectacular setback in the 2005 Dover, Pennsylvania, intelligent-design trial (*Nature* 439, 6–7; 2006), and the 2006 electoral losses in Ohio and Kansas, the movement could use some help — and Behe is eager to provide it.

Knowing how easy it is to demonstrate the workings of evolution in the development of drug resistance in viruses, bacteria and protozoan parasites, Behe concedes the point that evolution works very well at this level. His case study, repeated almost to the point of tedium, is malaria. Resistance to drugs such as chloroquine has indeed arisen within the parasite population, and so has genetic resistance to the parasite in humans. But if the inter-species genetic warfare between *Homo sapiens* and *Plasmodium* is actually a prime example of evolution, how can it then be used to make the case for 'design'?

The reason can be found in the book's title. To Behe, the genetic changes in both parasite and host represent the absolute limit of what darwinian processes can accomplish, and mark the "edge of evolution". He describes these well-understood mutations as a kind of "trench warfare" in which parasite and host endure a series of destructive mutations in key elements of cellular machinery. These changes produce nothing genuinely new, serving only to block the parasite or render widely used drugs ineffective by altering target proteins or clearing damaged cells. The fact that centuries of conflict between parasite and host have produced nothing in the way of new, complex systems in either species is proof that this is all that evolution can do. They mark the "limits of darwinism".

Where does this leave evolutionary explanations of more complex systems? Behe tells us frankly that darwinism cannot account for even a modest share of the complexity of life, and therefore design is absolutely required as an explanation. Yet, at the heart of his anti-darwinian calculus are numbers not merely incorrect, but so spectacularly wrong that this



Far from finding a clean edge to evolution, Behe's poorly designed arguments crumble away.

badly designed argument collapses under its own weight.

Behe cites the malaria literature to note that two amino-acid changes in the digestive-vacuole membrane protein PfCRT (at positions 76 and 220) of *Plasmodium* are required to confer chloroquine resistance. From a report that spontaneous resistance to the drug can be found in roughly 1 parasite in 10^{20} , he asserts that these are the odds of both mutations arising in a single organism, and uses them to make this sweeping assertion:

"On average, for humans to achieve a mutation like this by chance, we would need to wait a hundred million times ten million years. Since that is many times the age of the universe, it's reasonable to conclude the following: No mutation that is of the same complexity as chloroquine resistance in malaria arose by Darwinian evolution in the line leading to humans in the past ten million years."

Behe, incredibly, thinks he has determined the odds of a mutation "of the same complexity" occurring in the human line. He hasn't. What he has actually done is to determine the odds of

these two exact mutations occurring simultaneously at precisely the same position in exactly the same gene in a single individual. He then leads his unsuspecting readers to believe that this spurious calculation is a hard and fast statistical barrier to the accumulation of enough variation to drive darwinian evolution.

It would be difficult to imagine a more breathtaking abuse of statistical genetics.

Behe obtains his probabilities by considering each mutation as an independent event, ruling out any role for cumulative selection, and requiring evolution to achieve an exact, predetermined result. Not only are each of these conditions unrealistic, but they do not apply even in the case of his chosen example. First, he overlooks the existence of chloroquine-resistant strains of malaria lacking one of the mutations he claims to be essential (at position 220). This matters, because it shows that there are several mutational routes to effective drug resistance. Second, and more importantly, Behe waves away evidence suggesting that chloroquine resistance may be the result of sequential, not simultaneous, mutations (*Science* 298, 74–75; 2002), boosted by the so-called ARMD (accelerated resistance to multiple drugs) phenotype, which is itself drug induced.

L. CIAPONI/GETTY IMAGES

A mistake of this magnitude anywhere in a book on science is bad enough, but Behe has built his entire thesis on this error. Telling his readers that the production of so much as a single new protein-to-protein binding site is “beyond the edge of evolution”, he proclaims darwinian evolution to be a hopeless failure. Apparently he has not followed recent studies exploring the evolution of hormone-receptor complexes by sequential mutations (*Science* 312, 97–101; 2006), the ‘evolvability’ of new functions in existing proteins — studies on serum paraxonase (PON1) traced the evolution of several new catalytic functions (*Nature Genet.* 37, 73–76; 2005) — or the modular evolution of cellular signalling circuitry (*Annu. Rev. Biochem.* 75, 655–680; 2006). Instead, he tells his readers that there is just one explanation that “encompasses the cellular foundation of life as a whole.” That explanation, of course, is intelligent design.

The sad mistake at the logical centre of this book is eerily reminiscent of a similar claim in Behe’s 1996 book *Darwin’s Black Box*. In this work he claimed that complex biochemical systems have a property he called “irreducible complexity”. Irreducibly complex structures, such as the bacterial flagellum, could not have evolved because they lack any selectable function until all of their component parts are in place. As he wrote, “any precursor to an irreducibly complex system is by definition nonfunctional”, since every part of such a system had to be in place for natural selection to favour it. Therefore, such structures must have been designed. A nice argument, except for the annoying fact that it is wrong. Subsets of proteins nearly identical to those in the flagellum do indeed have selectable functions (*Nature Rev. Microbiol.* 4, 784–790; 2006), and the argument fails. In the same book, Behe also claimed that every component of the irreducibly complex vertebrate blood-clotting system had to be present for the system to work properly. That argument collapsed when Russell Doolittle’s laboratory (*Proc. Natl Acad. Sci.* 100, 7527–7532; 2003) showed that the puffer fish, *Fugu*, lacks at least three clotting factors and still has a workable system. Such failures in the science of the argument helped to send intelligent design to a defeat in the Dover trial, and they haunt it still.

No doubt creationists who long for a scientific champion will overlook the parts of this deeply flawed book that might trouble them, including Behe’s admission that “common descent is true”, and that our species shares a common ancestor with the chimpanzee. Instead, they will cling to Behe’s mistaken calculations, and proclaim that the end of evolution is at hand. What this book actually demonstrates, however, is the intellectual desperation of the intelligent-design movement as it struggles to survive in the absence of even a shred of scientific data in its favour. ■

Kenneth R. Miller is professor of biology at Brown University, Providence, Rhode Island 02912, USA.

EXHIBITION

A Wellcome addition

The Wellcome Collection

by the Wellcome Trust

183 Euston Road, London

www.wellcomecollection.org

Sara Abdulla

A unique cultural venue opened in London this month. The Wellcome Collection is the first permanent home for the massive, maverick history-of-medicine collection that pharmaceutical entrepreneur Sir Henry Wellcome (1853–1936) gathered throughout his life. Thirty million pounds (US\$60 million) and decades in the making, the free venue has three galleries, one of the world’s most important history-of-medicine libraries, an original programme of live events, a members’ club, a bookshop, a café, a conference centre and Pablo Picasso’s Bernal mural.

Wellcome’s fortune also created the Wellcome Trust, Britain’s main bioscience research funding agency. The trust has now remodelled the compact 1930s building it recently vacated to realize Sir Henry’s vision of a ‘Museum of Man’ and to extend its public engagement activities.

The scholarly heat rises with each floor. Street level lures in passers-by from the thundering road outside with a chic café and striking large-scale works — including a pendulous Antony Gormley cast. Here temporary exhibitions will explore the interplay between advances in medical science and our view of ourselves. The opening show, *The Heart*, runs until 14 September 2007; it features Andy Warhol prints, Leonardo da Vinci anatomical drawings and a wall of fixed animal hearts.

The next floor boasts two permanent galleries charting the evolution of our cultural response to health, sickness and discovery. *Medicine Man* displays some of Sir Henry’s extraordinary anthropological and ethnographic haul, such as these Chinese porcelain fruits containing couples engaged in sexual foreplay (pictured). *Medicine Man* has been seen in public just once before, at the British Museum in 2003 (*Nature* 423, 805; 2003). Looking like a cross between the Horniman Museum

and a Hollywood humidor, its handsome walnut cases, drawers and hidden cabinets reveal a telling fraction of the objects Sir Henry amassed. There are amputation saws, birthing tools, diagnostic dolls, arresting paintings and glassware galore. *Medicine Now* brings the story of ‘what it means to be human’ up to date, in a bright white and red journey through malaria, obesity, genomics and more.

Throughout the building, subtle curation and sumptuous display invite visitors to reflect on our knowledge, hopes, fears and beliefs about the body. This dialogue will continue in *The Forum*, an auditorium for debates, workshops, lectures and performances. Some of these will engage with themes of the temporary exhibitions. For example on 5 July, the audience can watch a live video link to a heart-valve operation, ask questions of the surgeon and examine instruments akin to those being used. Other events, such as the Islam and medicine panel on 19 July, will respond to current affairs.

The second floor brings the trust’s vast library into the twenty-first century. Virtual browsing stations and WiFi now complement the graceful galleries long beloved by science- and society scholars (and TV crews in search of instant gravitas). The top floors house The Wellcome Trust Centre for the History of Medicine at University College London, where much of this thoughtful activity starts.

And what of the members’ club? Will it become biology’s Algonquin Hotel? Quite possibly: it is inside a thrilling new museum, beside a leading medical school, opposite London’s new European rail terminus and encircled by scientific publishers. What better place to raise a glass to humane curiosity, the legacy of Sir Henry Wellcome. ■

Sara Abdulla is *Nature*’s chief commissioning editor.



WELLCOME LIBRARY, LONDON

Life drawing

Maria Sibylla Merian (1647–1717) was a woman ahead of her time. Born in Germany into a family of artists and engravers, she worked as an illustrator. She is best known for the paintings she made on a trip to Surinam in 1699. This was well before the more famous male naturalists/explorers such as Charles Darwin and Alfred Russel Wallace travelled the globe, bringing the natural diversity of far-off places to public attention.

Merian was a naturalist from childhood. She made great efforts to depict the life cycle of the animals she drew, often including

a key food plant in the picture. The frog illustrated here is from Merian's *Drawings of Surinam Insects*.

In his *Systema naturae*, the first taxonomic work describing and naming animal and plant species, Carl Linnaeus mentions Merian's work 136 times.

Two new books tell the story of this remarkable woman. In *Die Tierwelt der Maria Sibylla Merian* (Basilisken-Presse, €96), Katharina Schmidt-Loske analyses the drawings and watercolours attributed to Merian. Kim Todd's biography *Chrysalis* (I. B. Tauris, £16.99) brings the life and work of Merian to a wider audience.



TRUSTEES OF THE BRITISH MUSEUM

A better way of thinking?

The Soulful Science: What Economists Really Do and Why It Matters

by Diane Coyle

Princeton University Press: 2007. 288 pp. \$27.95/£17.95

Frances Cairncross

Alan Blinder, an American economist, once conducted a survey among his friends. He asked them whether there should be one queue in a crowded cafeteria at lunchtime, or two: a fast one for people paying higher prices and a slow one for everyone else. All the economists he knew said two; all the non-economists one. Diane Coyle recounts this telling anecdote in her ambitious book *The Soulful Science*.

Economists typically see the world through a different lens from most folk. "Economics isn't defined by its subject matter but by its way of thinking," says Coyle. But is it a better way of thinking? She sets out to argue that it is, and that economics has had an unfairly bad press. She comes to her subject from a background as a practitioner rather than as an academic: she has been an economics journalist, runs a consultancy, and has written several books with catchy titles, such as *The Weightless World* and *Sex, Drugs and Economics* that make new ideas in economics popular and accessible. Her writing style is clear but sometimes plodding, although she has a good line in anecdotes about the foibles of economists. She describes, for instance, how Joseph Stiglitz, invited to spend a term at the University of Oxford in exchange for eight hours of lecturing, spent one day on a marathon lecture and the rest of the term on his research. As one might say of much in economics — rational but infuriating.

Coyle's defence of her discipline is in some ways the most unsatisfactory aspect of her

book. Economists have headed down plenty of blind alleys, and policy-makers have frequently followed them. The first third of the book, devoted to explaining why countries grow and how economic development occurs, contains plenty of examples. One of the most telling chapters describes how economists first provided the justification for massive programmes of development aid — and are now assembling compelling evidence that official aid is either useless as a tool to promote growth or actually harmful.

The critics, Coyle would retort, are often out of date: economics is undergoing a revolution, and becoming a more useful and effective approach to the world. The discipline, she says, is at the threshold of a "golden age", in which many conventional ideas will be overturned. Two changes, little perceived in the outside world, are transforming the discipline. The more important is a revolution in the tools available to economists. We are seeing the creation of a vast number of large data sets, such as figures for historical and cross-country gross domestic product (GDP), and panel studies to track groups as diverse as ethnic-minority children or adults reaching retirement. We are also seeing the development of new statistical techniques and ways to scrutinize the details of economic activity that allow economists to examine more precisely the reasons people make particular choices. And powerful computing has revolutionized economics just as dramatically as other disciplines that rely on gathering and analysing vast quantities of data.

The second change — less important but more interesting to the ordinary reader — has been the willingness of economists to explore and borrow from other approaches to understanding human behaviour. Economists are

ransacking psychology, biology and even zoology. Experiments with capuchin monkeys suggest that conventional neoclassical assumptions about the impact of prices on demand apply even in parts of the animal kingdom, reports Coyle. Give monkeys tokens to 'buy' grapes or jelly, and a fall in the relative 'price' of grapes will cause them to buy less jelly and more fruit.

One limitation of such multidisciplinary approaches is that sociology and psychology often take a less rigorously numerate approach to the world, relying on qualities such as 'culture' that are hard to quantify. Economists are getting steadily better at measuring, but still tend to feel that what can't be measured isn't real. But what *can* be measured, and scrutinized through the unique lens of economics, reveals ever more fascinating perspectives on the world. Books such as *Freakonomics* by Steven Levitt and Stephen Dubner, or Tim Harford's 'Undercover Economist' column in the *Financial Times*, draw on the combined revolution in tools and perspectives that is transforming economics.

These arguments are not, however, the book's most impressive achievement. The best thing about it is a deft mapping of the developments in economic thought. Coyle describes brilliantly the intellectual geography of her subject. So anyone who wants to understand how the big ideas of the past half century fit together, and which economists' publications have been key in advancing particular arguments, could not do better than read this book. For a student wondering whether to study the subject, this is the perfect introduction. Although perhaps the basic test for a would-be economist is really very simple: would you prefer your approach to the lunch counter to be rationed by time or by price? One queue or two? ■

Frances Cairncross is rector of Exeter College, University of Oxford, Oxford OX1 3DP, UK, and a former journalist on *The Economist* magazine.

NATL. MARITIME MUS., LONDON



D. GILBERT, ©NATL. MARITIME MUS., LONDON

Unusual design:
new planetarium
embodies geometry
and astronomy.

The clever cone

A new planetarium for the Royal Observatory in Greenwich.

Martin Kemp

Traditionally planetariums have been housed in domed structures, echoing the 'vault of the heavens'. As we approach the new Peter Harrison Planetarium at Greenwich, London, across the terrace overlooking the Thames, high above Inigo Jones's Queen's House and Christopher Wren's Greenwich Hospital, our expectations are confounded. A strange mottled bronze cone emerges from the paving, like the protruding tail of an alien space craft that has crash-landed, deeply embedding itself in the ground.

As we walk around the cone, its form progressively declares itself. One side is vertical, whereas the opposite contour emerges at an angle from the ground. The cone has been sliced to produce an elliptical face, mirrored to reflect the passing sky.

It transpires that the inscribed angle of the inclined face is 51.5° , corresponding to the latitude of Greenwich. The conic section is made at 90° to this face, which means that the mirrored plane is parallel to the Equator. The vertical edge designates the zenith, while an inscribed line running up the sloping contour acts as a sighting line for the North Star.

This is an unquestionably clever building, which acts as visual and pedagogic compliment to the inscribed sphere of the 120-seat internal viewing chamber. It has been designed by architects Allies and

Morrison as part of the £15-million (US\$30-million) redevelopment of the historic complex of former Royal Observatory buildings.

The site is literally central to the history of the measurement of the world and of the cosmos, in keeping with the idea of "man as the measure of all things" (to quote the ancient tag credited to Protagoras). It was here in 1675, by the decree of Charles II, that Christopher Wren built the observatory for John Flamsteed, with the express purpose of finding an exact way to determine longitude. The quest, as we know, was to be prolonged, and only in 1773 was John Harrison's H4 timekeeper recognized as providing the means to solve the problem.

The worldwide adoption of Greenwich Mean Time as the international point of reference was enshrined in 1884 with the recognition at the Washington gathering of 25 nations that the Greenwich meridian should henceforth be regarded as longitude 0° .

Looking at the Peter Harrison Planetarium in its historic setting, resonances abound. Not least, it is elegantly in keeping with the modes of thought that animated Wren, Robert Hooke and other luminaries of the early Royal Society. The immanent geometry of nature, whether the mighty ellipses of the planets or the miniature 'engines' revealed by the microscope, was not only the fit subject of science but also

provided human designers with their basic vocabulary of form.

Conic sections and newly defined curves lay at the heart of Wren's work in architecture no less than in his science. He stressed that "the geometrical is the most essential part of architecture". By "geometrical" he meant not so much the abstract basis of beauty but the embeddedness of geometry in engineered structures, natural and artificial.

As an astronomer, student of dynamics and statics, meteorologist and inventor of scientific devices, Wren's practical extraction of natural design was perpetually at the heart of his endeavour. This understanding was expressed above in his domes, including the twin structures of Greenwich Hospital and, of course, that of St Paul's Cathedral.

Wren and his contemporaries would have been surprised by the actual form and materials of the new planetarium, which inevitably lie outside their architectural vocabulary of the seventeenth century, but they would have felt completely at home with the geometry and astronomy that it embodies. The new planetarium is a fitting addition to a site that demands architecture of a high order.

Martin Kemp is professor of the history of art at the University of Oxford, Oxford OX1 1PT, UK. His new book, *Seen | Unseen*, is published by Oxford University Press.

NEUROPHYSIOLOGY

Stressful pacemaking

Bruce P. Bean

In Parkinson's disease, dopamine-secreting neurons die — perhaps because unrelenting calcium entry during spontaneous electrical activity puts them under unusual pressure.

At a conference on calcium-channel-blocking drugs in the 1980s, a neurologist friend told me of colleagues who kept pills of the calcium-channel blocker nimodipine handy, planning to ingest one immediately should they suffer a stroke. Showing even more faith in the drug, an executive from the pharmaceutical company sponsoring the conference said that he took an unprescribed nimodipine pill every morning with his cereal, on the assumption that it was better not to wait.

Nimodipine and other dihydropyridine drugs were originally developed to treat high blood pressure; they relax vascular smooth muscle by blocking calcium entry. Subsequent off-label use to treat stroke was based partly on the idea that these drugs might also minimize neuronal death resulting from excessive calcium entry following oxygen deprivation. But despite the enthusiasm of my colleagues, controlled clinical trials¹ failed to show any beneficial effects of nimodipine for treating ischaemic stroke. Now, however, a study by Chan *et al.*² (page 1081 of this issue) has raised the exciting possibility that these or similar calcium-channel-blocking drugs might provide a strategy to treat Parkinson's disease by their effect on particular neurons — those that act by secreting the neurotransmitter dopamine (dopaminergic neurons).

The critical event in Parkinson's disease is death of dopaminergic neurons in a region of the brain known as the substantia nigra (SNc). Progressive loss of these neurons produces devastating symptoms, including tremors and loss of voluntary movement. Despite intensive research, it is not known what the causative pathological events are, nor why they selectively damage dopaminergic neurons. Moreover, current treatment cannot slow disease progression, only ameliorating its symptoms³.

Most commonly, activity in a neuron is triggered by neurotransmitter released by other neurons and diffusing across the gap, or synapse, between the neurons. But dopaminergic neurons in the SNc are spontaneously active even without synaptic input, firing action potentials at about 1–2 hertz (Fig. 1a). Such autonomous 'pacemaking' activity is seen in many types of neuron and requires ion channels that can open at membrane potentials lower

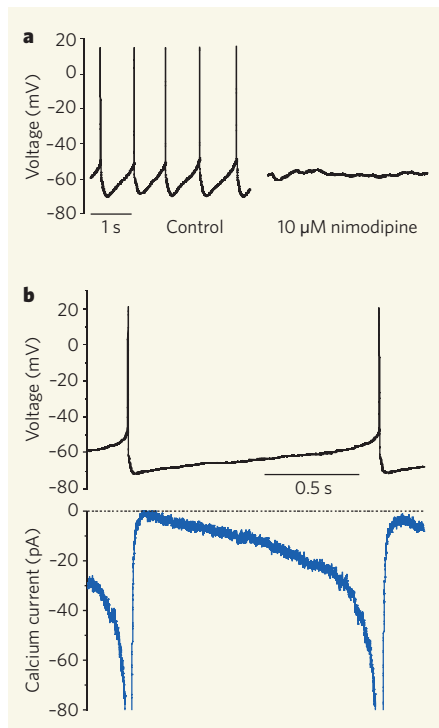


Figure 1 | Electrical pacemaking and calcium currents. **a**, Electrical activity in a neuron from the substantia nigra region of the brain shows rhythmic, spontaneous firing of action potentials. Application of the calcium-channel blocker nimodipine silences firing (unpublished recording by M. Puopolo). **b**, The time course of the calcium current during the pacemaking cycle was determined in a voltage-clamp experiment in which a record of spontaneous firing was used as the command voltage (thus controlling the cell's voltage to exactly match a natural pacemaking cycle). The calcium current was then determined when the current was blocked by substituting Co^{2+} for Ca^{2+} ions in the extracellular solution. Current in the negative direction corresponds to calcium ions entering the neuron. (Redrawn with permission from ref. 9.)

than the threshold for firing action potentials. The electrical current entering the cell through these channels then depolarizes the membrane to the threshold for action potentials.

In most pacemaking neurons, spontaneous sub-threshold depolarization is caused by

sodium entry, either through hyperpolarization-activated cation (I_h) channels, or through voltage-dependent sodium channels, or both. Neurons in the SNc are unusual in that the current driving pacemaking is carried mainly by calcium ions^{4,5}.

Chan and colleagues² began by testing for disruption of pacemaking in mice lacking a particular calcium channel, known as $\text{Ca}_v1.3$, which can open at sub-threshold voltages^{6–8}. Pacemaking in SNc neurons from these mice still occurred, but, in contrast to pacemaking in normal neurons, it could be stopped by blocking I_h or sodium channels, indicating a change in the underlying mechanism. By systematically examining the properties of I_h and sodium channels, the authors pinpointed this change to a shift in the voltage dependence of I_h channels that causes more current to flow at sub-threshold voltages.

The ability of SNc neurons to switch to a different mechanism of pacemaking with loss of $\text{Ca}_v1.3$ channels is an interesting example of adaptive plasticity to maintain physiological function. Unexpectedly, Chan and colleagues found that this plasticity does not require long-term changes, but can occur in neurons from normal mice within hours. When calcium channels were blocked by the dihydropyridine drug isradipine, cells initially stopped firing but, in the continued presence of blocker, they gradually resumed pacemaking, so that most cells were firing at near-normal frequencies within three to four hours. The resumed pacemaking was stopped by blocking I_h or sodium channels, as occurred in mice lacking $\text{Ca}_v1.3$ channels. The authors call the shift in firing mechanism 'rejuvenation', because the altered mechanism resembles pacemaking in young (less than 3 weeks old) normal mice, which also relies on I_h and sodium currents (a shift to the calcium-current-dependent mechanism occurs later, as mice mature).

Calcium can enter neurons through many pathways, but generally the period of entry is brief, as it is for entry through voltage-dependent calcium channels activated by action potentials or through synaptically activated NMDA-receptor channels. Although the calcium current underlying pacemaking in SNc

neurons is not large, it is almost continuously maintained⁹ (Fig. 1b). Furthermore, it occurs throughout the network of branching processes called dendrites that constitute most of the neurons' surface area⁵. Chan *et al.*² show that, at least in dendrites, pacemaking-associated calcium entry far outweighs that elicited during action potentials.

Chan *et al.* propose that this unusual, unrelenting calcium entry constitutes an energetic stress that makes SNc neurons particularly susceptible to oxidative damage and death. The uptake of calcium by cellular organelles called mitochondria — the powerhouses of the cell — is a crucial element in handling large loads of this ion^{10,11}. So it is plausible that continuous calcium entry would produce mitochondrial stress, which has been implicated in Parkinson's disease. Rare familial forms of the disease arise from mutated mitochondrial enzymes, and administration of mitochondrial poisons such as rotenone and MPTP provides a model of Parkinson's disease in rodents^{3,12}.

Consistent with their hypothesis that calcium entry through Ca_v1.3 channels stresses mitochondria, the authors show that rotenone has a far less damaging effect on SNc dopaminergic neurons in brain slices from Ca_v1.3-deficient mice than in those from normal mice, and that damage can be reduced by isradipine in normal neurons. They also show that, in a mouse model of Parkinson's disease produced by MPTP injection, isradipine administered *in vivo* reduced both the loss of SNc neurons and the development of movement disorders.

These are exciting results, because dihydropyridine calcium-channel blockers are well-studied drugs with relatively mild side effects. However, there are caveats. It is uncertain whether the 'rejuvenated' pacemaking in dihydropyridine-treated SNc neurons leads to adequate dopamine release, especially from dendrites. Moreover, the nimodipine concentrations used in the *in vitro* studies by Chan *et al.* were far higher than plasma concentrations of this drug that occur in its clinical use for treating high blood pressure, and thus might produce more severe side effects than are so far known.

On the other hand, nimodipine and isradipine were originally identified by their ability to block Ca_v1.2, not Ca_v1.3, channels, because they were assayed for relaxing strips of vascular smooth muscle where Ca_v1.2 is the dominant form. In fact, they are 10–20-fold more potent in blocking Ca_v1.2 than Ca_v1.3 channels^{7,8}.

Large libraries of potential calcium-channel blockers have been synthesized, and it is conceivable that other blockers selective for Ca_v1.3 channels already exist and can be identified by new screens. But even drugs that are highly selective for Ca_v1.3 channels might have unwanted side effects, such as slowing the heart-beat. This is seen in Ca_v1.3-deficient mice⁶, in which, surprisingly, adaptive homeostatic mechanisms seem less able to fully restore pacemaking in cardiac cells than in SNc dopaminergic

neurons. The Ca_v1.3 channel is also prominent in neuroendocrine cells and auditory hair cells⁶, so blocking these channels might have other potentially serious side effects.

Addressing the issue of side effects may turn out to be challenging. Nonetheless, by presenting a plausible hypothesis for why SNc neurons are selectively killed in Parkinson's disease, Chan and colleagues² have identified a promising target for developing drugs to slow the progression of this devastating disease. ■

Bruce P. Bean is in the Department of Neurobiology, Harvard Medical School, 220 Longwood Avenue, Boston,

Massachusetts 02115, USA.

e-mail: bruce_bean@hms.harvard.edu

1. Kaste, M. *et al.* *Stroke* **25**, 1348–1353 (1994).
2. Chan, C. S. *et al.* *Nature* **447**, 1081–1086 (2007).
3. Dauer, W. & Przedborski, S. *Neuron* **39**, 889–909 (2003).
4. Fujimura, K. & Matsuda, Y. *Neurosci. Lett.* **104**, 53–57 (1989).
5. Wilson, C. J. & Callaway, J. C. *J. Neurophysiol.* **83**, 3084–3100 (2000).
6. Platzer, J. *et al.* *Cell* **102**, 89–97 (2000).
7. Lipscombe, D. *et al.* *J. Neurophysiol.* **92**, 2633–2641 (2004).
8. Koschak, A. *et al.* *J. Biol. Chem.* **276**, 22100–22106 (2001).
9. Puopolo, M. *et al.* *J. Neurosci.* **27**, 645–656 (2007).
10. Herrington, J. *et al.* *Neuron* **16**, 219–228 (1996).
11. Patterson, M. *et al.* *J. Gen. Physiol.* **129**, 29–56 (2007).
12. Lin, M. T. & Beal, M. F. *Nature* **443**, 787–795 (2006).

EARTH SCIENCE

Silicon-enhanced core

Tim Elliott

What elements, besides iron, make up Earth's core? Discrepancies in the isotopic ratios found in rocks from Earth's mantle and in undisturbed meteoritic material indicate strongly that one answer is silicon.

Earth's core is its most inaccessible part. New information on its composition — such as that revealed by Georg *et al.* on page 1102 of this issue¹ — is therefore a hard-won prize. The core is known to be made predominantly of iron, but geophysical estimates of its density require that a lighter element makes up some 10% of its mass. Hydrogen, carbon, oxygen, sulphur and silicon have all been fingered as culprits, but which elements are actually involved remains controversial.

Despite what Hollywood would have us believe, we can't sample Earth's core directly. Instead, we must assess its composition by indirect means. One such method is mass balance: the budget of elements not in the silicate-dominated, outer portions of the Earth must be in the core (Fig. 1). For many elements of interest, the well-mixed, convecting mantle is the only significant reservoir in these outer layers. Thus, if we know the composition of the mantle and of the Earth as a whole, the make-up of the core can be calculated from the difference.

That might sound straightforward. Earth's bulk composition is estimated from analyses of 'undifferentiated' meteorites, thought to be made of the same primordial material from which Earth originally formed. And the composition of the mantle, normally hidden beneath a veneer of crust, can be determined from rare fragments fortuitously exposed at the surface. But significant chemical variability in both the meteorite and mantle samples makes compositional estimates of the bulk Earth and mantle uncertain — and constraints on the make-up of the core poorer still.

Georg *et al.*¹ investigate the possible presence of silicon in Earth's core using a novel isotopic, rather than elemental, mass balance. Different

isotopes of an element show the same chemical behaviour, but form bonds of slightly different strengths. This can result in mass differences (fractionations) between reactants and products in chemical reactions. Sometimes, such mass differences are very small. Georg *et al.* show that this is apparently so for silicon isotopes in silicate melts and minerals. They find that the silicon isotope ratios — expressed as $\delta^{30}\text{Si}$, or parts per thousand difference in the ratio $^{30}\text{Si}/^{28}\text{Si}$ relative to a reference standard — of a range of mantle-derived silicates are very similar. The silicon isotopic composition of the silicate portion of Earth is thus well defined.

Chemical reactions that involve a large change in the bond strength or coordination of silicon, on the other hand, can lead to significant isotopic differences². Georg and colleagues' first-principles calculations indicate that the partitioning of silicon between silicate melts characteristic of the mantle, and metallic melts such as are found in Earth's core, is such a process. If sufficient silicon is incorporated into the core, the residual silicon in the mantle should be isotopically heavier. Given the impressively high precision of the authors' analyses, this difference should be detectable.

The authors compare¹ the $\delta^{30}\text{Si}$ values of their mantle samples with those of a range of undifferentiated meteorites, which act as a reference for the bulk Earth. Strikingly, these meteorites show only minor variability in $\delta^{30}\text{Si}$, and are clearly resolved from the mantle samples, which are indeed isotopically heavier. That would seem to be smoking-gun evidence for silicon having been incorporated into Earth's core.

This conclusion resonates with other experimental^{3,4} and observational⁵ constraints. The

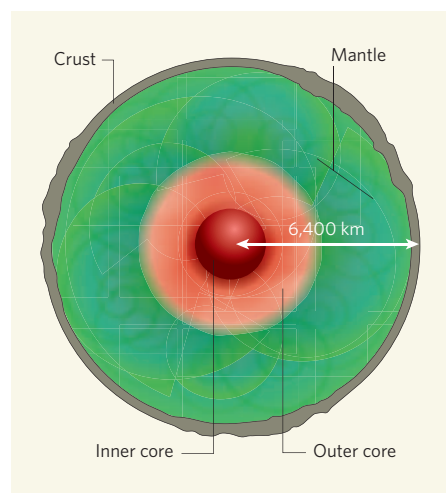


Figure 1 | Divided Earth. Earth consists of iron and silicate portions. The core is predominantly iron (with around 5% nickel and 10% of an as-yet unidentified additional 'light' element), and consists of a crystallized inner core and a molten outer core. The core segregated early in Earth's history from the overlying silicate portion. This is predominantly mantle, which is thought to convect and thus be well mixed (at least on length scales greater than around 10 km). The 6–60-km-thick veneer of crust covering the mantle is compositionally different, but is not a significant reservoir for the 'major' elements such as silicon, the element considered by Georg and colleagues¹. But it does hinder access to the mantle, which must be sampled either indirectly via melts that are erupted at volcanoes, or by rare fragments brought to the surface by tectonic activity.

high pressures associated with core formation on Earth would greatly enhance silicon concentrations there compared with those found on smaller planets such as Mars. In this way, Georg *et al.* can also explain why anomalously heavy $\delta^{30}\text{Si}$ values are not recorded in meteorites that represent the martian mantle.

The presence of silicon in Earth's core is significant because it requires chemically very reducing conditions during its formation. The corollary of this is that Earth's mantle must have subsequently evolved to its current more oxidized state³. This oxidation might result in fractionation of iron isotopes, and the authors propose this as an explanation for the anomalously heavy iron isotopic signature that is also a characteristic of Earth's mantle⁶. They thus present an elegantly self-consistent model of the isotopic consequences of core segregation in Earth's early history.

This model lays down the gauntlet for experimental verification. Stable-isotope fractionations diminish with increasing temperature², and so the large differences in silicon isotopes predicted by the authors for core-formation temperatures of around 2,000 kelvin might initially seem surprising. Although the estimates are consistent with the sense and magnitude required to account for their isotopic data, there are currently no high-temperature experimental data to calibrate the calculations.

Such data will be necessary, not least to obtain determinations of silicon-isotope fractionation that are accurate enough to try to quantify the amount of silicon in Earth's core.

There are, by contrast, some existing experimental data on high-temperature iron-isotope fractionations⁷. The inference from these data is that oxidation processes are not sufficient to account for the isotopically heavy iron in Earth's mantle (see Georg and colleagues' Supplementary Information). But here additional factors, such as changes in iron's spin state at high pressures, which affect bond strengths and so fractionation, need to be explored experimentally.

Clearly, new studies on both silicon- and iron-isotope fractionation are required to resolve outstanding uncertainties about Earth's

composition. The findings of Georg *et al.*¹ will give further impetus to this fundamental area of research.

Tim Elliott is in the Department of Earth Sciences, University of Bristol, Wills Memorial Building, Queen's Road, Bristol BS8 1RJ, UK.
e-mail: tim.elliott@bristol.ac.uk

1. Georg, R. B., Halliday, A. N., Schauble, E. A. & Reynolds, B. C. *Nature* **447**, 1102–1106 (2007).
2. Bigeleisen, J. & Mayer, M. G. *J. Chem. Phys.* **15**, 261–267 (1947).
3. Wade, J. & Wood, B. J. *Earth Planet. Sci. Lett.* **236**, 78–95 (2005).
4. Takafuji, N., Hirose, K., Mitome, M. & Bando, Y. *Geophys. Res. Lett.* **32**, doi:10.1029/2005GL022773 (2005).
5. Hart, S. R. & Zindler, A. *Chem. Geol.* **57**, 247–267 (1986).
6. Poitrasson, F., Levasseur, S. & Teutsch, N. *Earth Planet. Sci. Lett.* **234**, 151–164 (2005).
7. Polyakov, V. B. & Mineev, S. D. *Geochim. Cosmochim. Acta* **64**, 849–865 (2000).

NEUROBIOLOGY

The currency of guessing

Paul Cisek

Vanilla or chocolate? Fight or flight? A career in academia or in industry? The neural processes of probabilistic decision-making provide clues about the 'common currency' through which decisions are made.

Our lives are filled with decisions, ranging from selecting a dessert to choosing a career. In many such situations, we make the decision by weighing the pros and cons for each of the options and selecting the one whose potential benefits most outweigh its costs. But how does the brain go about comparing costs and benefits?

Neurophysiological studies in monkeys are providing insights into the neural mechanisms that underlie this process^{1–5}. So far, these studies have looked only at the simplest kinds of perceptual judgements or simple guesses, but the principles they reveal may apply even to the most sophisticated decisions of which humans are capable. Yang and Shadlen (page 1075 of this issue)⁶ now describe an investigation of how monkeys carry out a rudimentary form of reasoning called probabilistic inference. The authors' goal was to find out how the brain combines evidence from multiple sources to make the best informed guess about the correct choice.

The task used by Yang and Shadlen consists of a set of trials in which a monkey performs an eye movement to one of two visual targets, one green and one red. In each trial, only one of the targets will be associated with a reward, and the monkey has to guess which one that is on the basis of a set of symbolic cues. These cues are four shapes randomly selected from a set of ten, presented one-by-one in the centre of the visual display. Each shape conveys a specific amount of information, in favour of either the red target or the green target. At the end of the

trial, after four shapes have been shown, the correct target is determined probabilistically on the basis of the sum of this information. The monkeys have to combine the evidence provided by each of the cues to select the target that is most likely to be correct. Ultimately, their choice is merely a guess, but it can be an optimal one if they have correctly learned the partial information conveyed by each symbol and combine the pieces appropriately.

It has been suggested that the optimal way to combine cues that provide only partial information about the appropriateness of a choice is to make computations based on the likelihood that a given cue will appear in situations where that choice is the correct one (Box 1, overleaf). That is, suppose that a triangle occurs in 50% of the trials in which the green target is correct but in only 5% of the trials in which the red target is correct. You should therefore treat triangles as a cue that favours the green target. Suppose that a square occurs in 20% of green trials and in 40% of red trials. This favours the red target, but not as strongly as the triangle favoured the green. Thus, the simultaneous presence of both a triangle and a square favours green more than red, and guessing 'green' is optimal.

In fact, the best way to combine these separate cues is to compute something called the 'logarithm of the likelihood ratio' or 'logLR' (Box 1)². Remarkably, the monkeys' behaviour clearly showed that they were quite accurate in estimating the best choice given a set of symbolic cues. This reveals that they were able to assess the evidence provided by each



50 YEARS AGO

On May 13, Prof. J. B. S. Haldane delivered at University College, London, a lecture in celebration of the centenary of Karl Pearson's birth... Haldane began by examining some criticisms, fair and unfair, of Pearson's work. All power corrupts, and Pearson's power as a head of an important department and the editor of an important journal led him to make mistakes... It has even been claimed that by espousing a false theory of heredity he was a hindrance to progress. Haldane believes, in fact, that Pearson's theory of heredity was incorrect. But so, he remarked, was Columbus's theory of geography. In his search for a self-consistent theory of evolution, Pearson devised methods which are not only indispensable in any discussion of evolution; they are essential in every serious application of statistics.

From *Nature* 29 June 1957.

100 YEARS AGO

A correspondent recently directed our attention to a sensational report that certain signals are regularly received at one of the Marconi wireless telegraph stations, and are believed to be communications from Mars or another planet. As Mars will be in opposition on July 6, and is well situated for observation in southern observatories, the rumour will probably be extensively circulated during the next few months. A copy of the report was sent, therefore, to Mr. Marconi, who has favoured us with the following reply:—"There is no truth whatever in the statement which has been freely published for the last year or two that mysterious signals have been received at Cape Clear from probably some distant planet. There is, in the first place, no wireless telegraph station at Cape Clear. The stray or vagrant electrical effects which do manifest themselves from time to time at wireless telegraph stations are due to atmospheric discharges or other natural causes..."

From *Nature* 27 June 1907.

Box 1 | Probabilistic inference

What is the best way to make an informed guess? Suppose you need to decide whether to take your umbrella with you on your way to the store. If on average it rains 10% of the time in your home town, you could simply decide to take a chance that it will be dry.

But you can do better than that if you first collect some extra information. Looking outside, you see some dark clouds in the sky. In your experience, such clouds are present 90% of the time that it rains, but also 30% of the time that it doesn't. You can also look at the barometer

(on the off chance that you actually own one). Suppose that the barometer reading is one that occurs on 40% of rainy days and 10% of dry days. Does this mean there will be rain?

The optimal approach is to add up the logarithms of the ratios of probabilities conveyed by each of these three pieces of information, and to guess that it will rain if the sum is greater than zero.

Each of the stimulus cues (dark clouds, barometer reading) conveys information proportional to the logarithm of the ratio of the likelihood of

seeing that stimulus if it rains divided by the likelihood of seeing it if it doesn't rain. This value is the logarithm of the likelihood ratio. The clouds give you a ratio of 90/30, whose logarithm (base 10) is about 0.48. The barometer gives you a ratio of 40/10, whose logarithm is 0.60. Finally, the overall chance of rain gives you a ratio of 10/90, whose logarithm is -0.95. The sum of all of these factors is 0.13. Because this is greater than zero, then the final verdict is that it's more likely that it will rain. So take your umbrella.

P.C.

shape and to combine it in an appropriate way — that is, even monkeys are capable of probabilistic inference.

Even more remarkable, however, was what Yang and Shadlen found when they examined neural activity during the task. They recorded neurons in a part of the cerebral cortex called the lateral intraparietal area (LIP). Neurons in this region respond to visual targets and are activated before eye movements occur, and that activity is modulated by a variety of decision factors^{1,4,7}. Yang and Shadlen reasoned that many of these observations can be explained if the activity of LIP neurons is related to logLR, the putative 'common currency' for probabilistic inference.

Indeed, when they plotted neural activity against logLR, they found a remarkable match. As logLR increased, so did neural activity. This relationship between neural firing rates and logLR was consistent across time in the trial, as new information was provided by the new symbols that appeared. That is, whenever a new symbol appeared, the neural activity changed as a function of the information provided by that symbol. Thus, the total neural activity reflected the cumulative estimate of logLR in a manner that predicted the monkeys' behavioural choices at the end of each trial.

Although the activity of LIP neurons co-varies with logLR, this is not exclusively what it encodes. The LIP is part of the sensorimotor system for guiding eye movements, and is a half-way house between the processes of visual target detection and movement implementation. Neural activity in the LIP has been shown to reflect the processes of visual attention⁸ and motor intention⁹, as well as many variables related to the decision, such as the probability of movement and the size of reward¹, recent pay-offs⁴ and relative desirability⁷. The fact that all of these variables seem to coexist within a structure used to guide movement may seem surprising at first, but it is a trend found in many other systems^{3,5,10}. Indeed, it seems that decisions about actions are made within the

same brain circuits responsible for planning and controlling those actions. Decision variables such as logLR seem to modulate neural activity in a large network distributed throughout the brain, as if their role is to push and pull our motor system towards one or other possible actions.

Yang and Shadlen's research⁶ builds on previous studies of simple discrimination and decision tasks, and takes a step towards understanding the more complex kinds of reasoning of which animals are capable. Monkeys can do a lot more than is usually asked of them during neurophysiological experiments. They (and we) can use many sources of information to make choices. Such sources include sensory cues that provide partial information about the state of the world, the subject's own beliefs about what that information means, and the probabilities of different outcomes as well as their costs and benefits.

The authors' results suggest that the common currency with which the brain constructs an informed guess is the log likelihood ratio. Their findings shed light on the neural mechanisms of probabilistic inference, and may ultimately help in understanding higher cognitive abilities such as reasoning and deliberation. ■ Paul Cisek is in the Department of Physiology, Université de Montréal, CP 6128 Succursale Centre-ville, Montréal, Québec H3C 3J7, Canada. e-mail: paul.cisek@umontreal.ca

1. Platt, M. L. & Glimcher, P. W. *Nature* **400**, 233–238 (1999).
2. Gold, J. I. & Shadlen, M. N. *Trends Cogn. Sci.* **5**, 10–16 (2001).
3. Romo, R., Hernandez, A. & Zainos, A. *Neuron* **41**, 165–173 (2004).
4. Sugrue, L. P., Corrado, G. S. & Newsome, W. T. *Science* **304**, 1782–1787 (2004).
5. Coe, B., Tomihara, K., Matsuzawa, M. & Hikosaka, O. *J. Neurosci.* **22**, 5081–5090 (2002).
6. Yang, T. & Shadlen, M. N. *Nature* **447**, 1075–1080 (2007).
7. Dorris, M. C. & Glimcher, P. W. *Neuron* **44**, 365–378 (2004).
8. Colby, C. L. & Goldberg, M. E. *Annu. Rev. Neurosci.* **22**, 319–349 (1999).
9. Snyder, L. H., Batista, A. P. & Andersen, R. A. *Vision Res.* **40**, 1433–1441 (2000).
10. Cisek, P. & Kalaska, J. F. *Neuron* **45**, 801–814 (2005).

ASTRONOMY

A new molecular factory

Sun Kwok

Carbon-rich stars are known to be prolific producers of molecules. Against expectations, astronomers have identified an old, oxygen-rich star that can also synthesize a chemically varied molecular cocktail.

For 30 years, astronomers' favourite star in the search for molecules was a nearby, old, carbon-rich star called CW Leo. This star produces such a variety of molecules that it constitutes a veritable 'molecular factory'. On page 1094 of this issue, Ziurys *et al.*¹ describe how they have discovered another stellar molecular factory that is almost as prolific. VY Canis Majoris (VY CMA) is a cool, old star like CW Leo, and like CW Leo it is in our Galactic neighbourhood. But it is different in one crucial respect: it is rich in oxygen. Apparently, then, oxygen-rich stars can also produce a wide variety of molecules when they are near the ends of their lives.

Most people think of stars as sources of light. But since the early 1970s, and the development of infrared and microwave astronomy, stars have also proved to be sources of matter. We now know that old stars generate strong stellar winds that eject material from the parent star into interstellar space. Because many heavy elements are synthesized by nuclear processes in the interior of old stars, stellar winds are a major avenue for feeding those elements into the interstellar medium.

What kinds of matter are in the stellar winds? In addition to atomic nuclei, it turns out that the stellar winds have very rich chemical contents. Spectroscopic observations at millimetre and submillimetre wavelengths have resulted in the detection of more than 60 types of molecule. These include inorganic molecules (for example, CO, SiO, SiS, NH₃, AlCl), organic molecules (C₂H₂, CH₄, H₂CO, CH₃CN), radicals (CN, C₂H, C₃, HCO⁺), rings (C₃H₂) and chains (HC₅N, HC₇N)². The winds are dispersing fast, so the formation processes must take place on timescales as short as hundreds of years.

Conventional wisdom suggests that if oxygen is more abundant than carbon, most of the carbon atoms in the wind of VY CMA should be locked up in the CO molecule, a simple and stable molecular species that forms in the atmosphere of the star. Given the lack of available carbon atoms, an oxygen-rich star is not expected to have the chemical versatility of a carbon-rich star and is therefore much less likely to have a chemically rich molecular content.

So it comes as a surprise that Ziurys *et al.*¹ announce seven newly discovered molecules in the stellar wind of VY CMA. Not only is the production of these increased numbers of molecules a feat in itself, but the variety

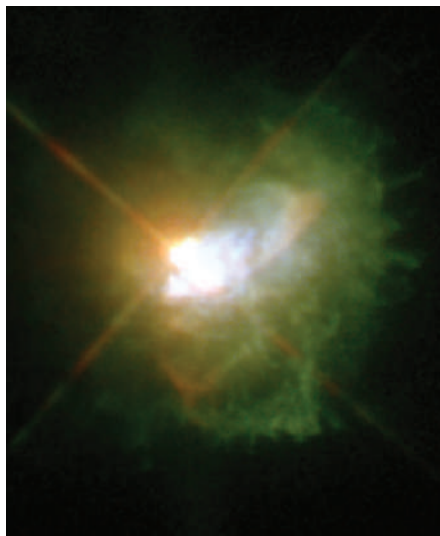


Figure 1 | Stellar performance. This colour-composite picture of VY CMA consists of separate images, taken in blue, green, red and near-infrared light, obtained with the Hubble Space Telescope (HST) Wide Field and Planetary Camera 2. The image reveals complex circumstellar ejecta, with arcs, filaments and knots of material formed by the massive outflows created during several outbursts. As Ziurys *et al.*¹ recount, the observed emission from different molecular species can be traced to the different optical features. (Composite image created from HST data taken by R. Humphreys.)

of molecules is striking as well. Among the seven are molecules containing sulphur (NS), phosphorus (PN), silicon (SiS) and even metals (NaCl, or salt). Phosphorus is of particular interest because this element has an essential role in the biochemistry of life, in spite of its low abundance in the cosmos. The authors' discovery also reveals unexpected physical complexity. They found the molecular ion HCO⁺, even though it is not expected to be present in the low-temperature, neutral environment of a cool star.

The molecules in VY CMA were identified through their rotational transitions. When a molecule decays to a lower rotational state, having been excited by collisions or radiation to a higher rotational state, it emits microwaves at specific wavelengths. Modern spectrometers can have high spectral resolution, so different molecular species can be uniquely identified by the wavelengths of the observed spectral lines. Because the observed line wavelengths are shifted by the velocities of the emitting gas,

as a result of the Doppler effect, Ziurys *et al.* were also able to derive the kinematic structures of the different components of the winds from the emission profiles of lines. Imaging of VY CMA by the Hubble Space Telescope has shown that the winds have complicated morphologies (Fig. 1), and the origin of the different molecules can be traced to different morphological features³. So these observations can also be used as probes to help elucidate the physics of the winds.

Our interest in old stars is not confined to the molecules they may be producing. In the gas phase, molecules are in turn believed to be the building-blocks of more complex inorganic and organic solids that consist of hundreds of atoms. These solids have been detected by infrared spectroscopy through their vibrational modes. Planetary nebulae, the descendants of old red-giant stars such as CW Leo, have been seen to contain carbonaceous compounds with both ring and chain structures typical of complex organic substances⁴. The detection of additional gas-phase molecules will help us draw a more complete picture of chemical synthesis in stars.

Ziurys and colleagues' report¹ shows that there is much more to be learnt about the use of old stars as laboratories for studying interstellar chemistry. How do these stars manage to make complex molecules in the extreme low-density environment? Do physical shocks, radiation or even grain-surface chemistry contribute to the molecular-formation process? What chemical pathways lead from the simple gas-phase molecules to complex organic solids? And what roles do these molecules have in the enrichment of the Galaxy and maybe even the early Solar System? Such questions have broad implications beyond astronomy and chemistry — winds from old stars will be an enticing target of observations for some time to come.

Sun Kwok is in the Department of Physics, University of Hong Kong, Pokfulam, Hong Kong. e-mail: sunkwok@hku.hk

1. Ziurys, L. M., Milam, S. N., Apponi, A. J. & Woolf, N. J. *Nature* **447**, 1094–1097 (2007).
2. Olofsson, H. in *Molecules in Astrophysics: Probes and Processes* IAU Symp. 178 (ed. Van Dishoeck, E.) 457–468 (Kluwer, Dordrecht, 1997).
3. Humphreys, R. M., Helton, L. A. & Jones, T. J. *Astron. J.* **133**, 2716–2729 (2007).
4. Kwok, S. *Nature* **430**, 985–991 (2004).

Robert Cahn (1924–2007)

Robert Cahn, who died on 9 April 2007, was a much-valued and prolific contributor to *Nature*, and to News & Views in particular. His articles first appeared in the mid-1960s, and ranged widely across the discipline of materials science. The July issue of *Nature Materials* includes a full obituary by A. Lindsay Greer, which is freely available at <http://dx.doi.org/10.1038/nmat1941>

Editor, News & Views

OBITUARY

Bohdan Paczyński (1940–2007)

Guiding light in astrophysics.

Bohdan Paczyński, one of the most brilliant astrophysicists of the twentieth century, died on 19 April 2007. Paczyński had an uncanny ability to see to the heart of complex astronomical phenomena, quickly identifying the essential physical processes. He performed this trick again and again, amply demonstrating that what he (with characteristic self-deprecation) called “common sense” was uncommon indeed.

Paczyński was born on 8 February 1940 in Wilno, Poland (now Vilnius, Lithuania). He was interested in astronomy from an early age, and as a teenager made many visual observations with a 25-centimetre refractor telescope to time the light minima of eclipsing binary stars. He went on to study physics at the University of Warsaw, receiving his PhD in 1964, after which he was based at the Copernicus Astronomical Center in Warsaw.

During the 1960s and through to the early 1980s, he was frequently invited to the West, and during one such visit — in December 1981 — martial law was declared in Poland. Paczyński chose to remain in the United States. In 1982 he was appointed to a professorship at Princeton University, where he spent the rest of his career.

Unlike our Sun, some stars vary considerably in their luminosity. To understand variable stars, Paczyński undertook pioneering studies on stellar structure and evolution; on the physics of the accretion disks of material that orbit many celestial bodies, and that mediate mass transfer in certain binary star systems; and on the evolutionary effects of this mass transfer. By 1967 he had published a series of theoretical papers on the evolution of close binary stars, and was recognized as a leading authority in this field.

Among his contributions were the first recognition that gravitational radiation influences the evolution of very tight binaries, and that there is a close relationship between the luminosity of a giant star and the mass of its degenerate core of helium ‘ash’. The Paczyński code for stellar evolution, one of the earliest examples of open-source scientific software, set a standard for the generous dissemination of research tools.

Paczyński was a leader in the field of gravitational lensing: the phenomenon whereby light from a source is bent by the gravity of an intervening body, such as a galaxy. His early papers on lensing by quasars and elliptical galaxies preceded the first observation of gravitational lensing by four years. In 1986, he proposed that modern

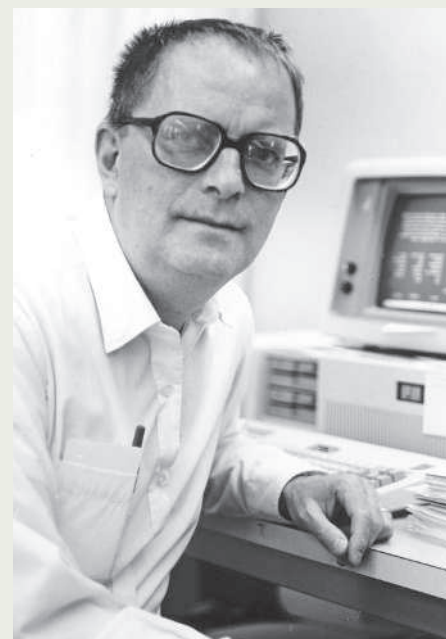
electronic detectors and computers could be used to search for rare gravitational ‘microlensing’ by stars or other compact masses in our own Galaxy. Although many astronomers thought this proposal wildly ambitious, his paper stimulated several groups — including his own Optical Gravitational Lensing Experiment (OGLE) team — to design and build suitable camera systems. This equipment allowed the observation in the 1990s of microlensing of stars in our Galaxy’s central bulge. The OGLE team has since detected thousands of microlensing events, providing strong evidence that ordinary matter dominates ‘dark matter’ in the Galactic bulge.

With Shude Mao, a graduate student, Paczyński went on to show that stellar companions, including planets, could have distinctive effects on microlensing light curves, allowing the mass and location of the companion to be determined. This paper led to the discovery of several extrasolar planets, including the first detection of an Earth-like planet.

In 1987, Paczyński was the first to suggest that ‘giant luminous arcs’ in two clusters of galaxies were in fact images of background galaxies, magnified and distorted by the gravitational field of the lensing cluster. Such cluster lens systems have since been used to observe distant galaxies that would otherwise be too faint to study, and the method provides an independent estimate of the masses of the clusters.

Paczyński also profoundly influenced the study of cosmic γ -ray bursts (GRBs) — hugely energetic flashes of γ -rays that are often followed by ‘afterglows’ at longer wavelengths. In 1986, he pointed out that observed GRBs had properties consistent with their being at cosmological distances, contradicting the prevailing view that GRBs originate on neutron stars within our Galaxy. Argument over the Galactic or cosmological origin of GRBs was heated, culminating in a public debate in Washington DC in April 1995 between Paczyński and Donald Lamb, moderated by Martin Rees. Within two years, Paczyński’s position was vindicated by observations of GRB afterglows, with absorption lines from intervening galaxies proving them to be at cosmological distances.

Variable stars, dwarf novae, novae, supernovae, GRBs — Paczyński was interested in, and delighted by, all cosmic phenomena and all of astrophysics. He suspected that there might be new kinds of as-yet undiscovered celestial events. He



R. P. MATTHEWS, PRINCETON UNIV.

understood that small telescopes had the potential to discover rare transient events, and with that in mind was responsible for the creation of the All Sky Automated Survey to systematically monitor an area of the sky containing tens of millions of stars.

Paczyński was always open to the possibility that fashionable models might be incorrect, and stressed the value of applying simple geometric or astronomical arguments whenever possible. His originality, coupled with his profound understanding of stars, gravitational lensing and, indeed, the essentials of all astrophysics, made him a unique source of new ideas. His achievements were recognized worldwide, and he received many honours, including the highest awards of the American Astronomical Society and the Royal Astronomical Society.

Bohdan Paczyński approached life with zest and good humour. One of his daily pleasures was discussing with his colleagues the latest preprints at morning coffee. He was very generous with his ideas; many a paper closes with thanks to Paczyński for valuable suggestions.

During the last four years of his life he battled brain cancer with courage and good cheer, continuing to do science to his last days. The world has lost a great scientist, and a wonderful man.

Bruce T. Draine and Jeremiah P. Ostriker

Bruce T. Draine and Jeremiah P. Ostriker are in the Department of Astrophysical Sciences, Princeton University, Princeton, New Jersey 08544-1001, USA.

e-mail: draine@astro.princeton.edu

Q&A

MATERIALS SCIENCE

Nanotube composites

Pulickel M. Ajayan and James M. Tour

A carbon revolution has occurred — carbon atoms can be coaxed into several topologies to make materials with unique properties. Nanotubes are the vanguard of this innovation, and are on the cusp of commercial exploitation as the multifunctional components of the next generation of composite materials.

Why do we need composite materials?

Composites, which are formed from two or more distinct materials, have desirable combinations of properties that aren't found in the individual components. For example, fibre-reinforced composites are engineered structures that commonly comprise a soft matrix — typically a polymer — encapsulating a stiffer, load-bearing filler in the form of fibres or particles. For high-strength composites, the fibres should be stiff and have a high aspect ratio (length-to-width ratio). This allows a good transfer of load from the matrix to the filler when the composite is put under mechanical stress, in much the same way that steel bars reinforce concrete.

What makes carbon nanotubes special compared with other reinforcing fibres?

It's the combination of their small size and particular physical properties. Nanotubes have molecular dimensions, and consist of perfect graphite sheets rolled into hollow cylinders. There are two types: single-walled (with diameters of about 0.5–2 nanometres) and multiple-walled (which have diameters of 2–50 nm). In many respects they resemble the polymer chains used as composite matrices — both have covalently bonded structures, similar dimensions and mechanical flexibility. This makes nanotubes entirely different from traditional fibres such as carbon or glass, which are relatively large (with diameters on the scale of micrometres) and brittle. The basic mechanical properties (such as strength) of nanotubes greatly exceed those of other fibres, yet this strength is combined with a low mass density, making them extremely light. Furthermore, nanotubes can conduct heat and electricity down their long axes as efficiently as metals — a great improvement on traditional fibres (Box 1).

How do nanotubes affect the properties of composites?

Nanotubes in plastic composites — which are anticipated to be the largest bulk application of these reinforcers — would serve to increase

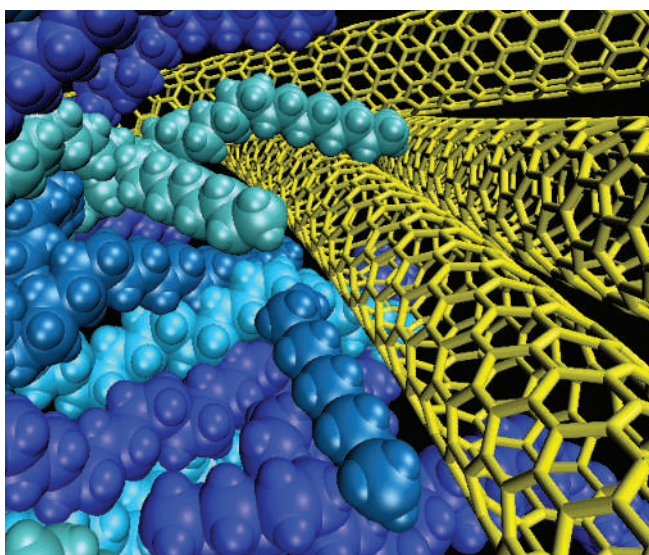


Figure 1 | Slippery interfaces. This computer-generated image shows how single-walled nanotubes (yellow) bundle together when used as the reinforcing element of a composite material. The nanotubes are depicted at the interface with the polymer polyethylene (individual polymer molecules are shown in different shades of blue). The surfaces of nanotubes are atomically smooth, so their interactions with the polymer matrix are weak.

P. SHEMELLA & C. SOLDANO

stiffness, strength and toughness, and provide other properties such as electrical and thermal conductivity. Because, at present, nanotubes can be manufactured only at lengths up to the submillimetre scale (so falling into the short-fibre category), for the foreseeable future their dominant role in composites is likely to remain as matrix modifiers and providers of multifunctional attributes. But once nanotubes can be efficiently assembled on a macroscopic scale, they could become serious competition for the continuous carbon fibres that are woven and stacked to form load-bearing elements in structural composites used in the building and engineering industries.

Are nanotube composites easy to make?

Unfortunately not. The biggest challenge is to fully disperse individual nanotubes in the matrices, because nanotubes tend to form clusters and bundles. These aggregates persist unless high shear forces are applied, for example by vigorous mixing of the polymer. But such mixing often damages nanotube structures, compromising their properties. Another issue is that the polymer–nanotube mixtures are highly viscous (owing to the large surface area of nanotubes). This creates engineering

problems, because the composites don't flow easily and are difficult to mould. Processing is also rendered difficult by the poor solubility of nanotubes in most solvents and polymers. Nevertheless, several approaches have been successfully adopted to obtain intimate mixing of nanotubes with polymer phases, including dry powder mixing, melt mixing, polymerization of monomers onto nanotube surfaces, and surfactant-assisted mixing.

Does a nanotube's small size directly affect the properties of nanotube composites?

Certainly. The reinforcement effect of a filler should improve markedly as its fibres get smaller, because of the increased surface area (per volume of filler) available for interaction with the matrix. For example, nanotubes of about 10 nm in diameter have an interfacial area with matrices that is almost 1,000 times greater than that of 10-micrometre fibres of the same volume. This intimate contact can influence the properties of the polymer matrix, affecting its chain entanglements, its morphology and even its crystallinity. Therefore, the resulting composite can be much tougher and stronger than would be expected

Box 1 | Nanotubes — a unique reinforcer

Unlike traditional micrometre-sized reinforcements, such as carbon fibres, nanotubes are truly molecular in size — comparable in lateral dimensions and aspect ratio to polymer chains. But there the comparison ends. In most polymers, cleavage of one bond is sufficient to sever the polymer strand. But nanotubes can require 10–50 carbon-carbon bonds per repeat unit to be broken for rupture to occur. As the carbon-carbon bond is one of the strongest found in nature, it is unlikely that there will ever be a polymer chain stronger than a nanotube.

The deformation characteristics of nanotubes are also intriguing. Carbon fibres are stiff and brittle, and act as rigid rods, whereas nanotubes are elastic and can be easily bent, buckled, twisted, flattened or tangled without breaking (see panel). So although nanotubes have much higher stiffness and tensile strength than carbon fibres, they are much more flexible: the bending of carbon fibres in composites leads to their



fracture, whereas long nanotubes bow, remaining in arbitrary shapes.

This difference results from the way that nanotubes and carbon fibres are formed from two-dimensional sheets of graphite (known as graphene layers). A single-walled nanotube can be thought of as an individual graphene sheet rolled up to make a defect-free, hollow structure just one atom thick — distinctly different from the randomly stacked graphene layers found in carbon fibres. The skinny carbon-shell structure of nanotubes makes

them strong in tension but easily bent; enables them to withstand large strains before failing; and provides a super-low-density framework.

The combination of high local stiffness with elasticity at larger scales determines the 'persistence length' of nanotubes — the characteristic distance from one end of the tube at which thermal energy becomes sufficient to introduce significant bending. The persistence lengths of single-walled nanotubes are expected to be in the order of tens to hundreds

of micrometres. For multiwalled nanotubes, or bundles of single-walled tubes, the persistence lengths could be substantially higher, becoming comparable to the maximum lengths of nanotubes that can currently be prepared. In these cases, the nanotubes behave like short, stiff carbon fibres. The load-bearing and conducting properties of nanotubes are bolstered by the unusual persistence lengths of these structures and by the varied, stable conformations that they adopt.

P.M.A. & J.M.T.

on the basis of the properties of the isolated filler. Furthermore, nanotubes with aspect ratios of tens of thousands to one can percolate — that is, form an inter-nanotube connective pathway — throughout the matrix at exceedingly small volume fractions, strengthening the composite and making it electrically conductive.

Why aren't nanotube composites already widely available?

Because there are still a few manufacturing problems to be overcome. Initial attempts to make these materials often yielded lacklustre results, mostly because of the poor interfaces that formed between the nanotubes and the surrounding matrix (Fig. 1) — strong bridges between the polymer and the nanotubes are necessary for good load transfer between the two. Such difficulties are not new, and also arose in the early days of carbon-fibre composites. However, the problem is more acute for nanotubes, both because their surfaces are atomically smooth (which results in poor matrix adhesion) and because the interfacial area for such tiny filler particles is huge.

Are there any other manufacturing issues?

Poor dispersion of nanotubes in the matrix is also a problem — large aggregates of nanotubes initiate cracks in composites, rather than reinforcing them. Bundle structures — comprising many parallel nanotubes held together

by weak bonds (Fig. 1) — also tend to form. Under stress, the nanotubes in the bundles can slip past each other, splitting off in layers from the host and weakening the composite. Examining the fracture surfaces of polymer-nanotube composites can help determine whether proper load transfer is occurring in the material (Fig. 2). If the matrix-filler interface is strong, the nanotubes break under stress. But for poorly fastened interfaces, nanotubes are pulled out of the matrix as a result of interfacial slippage.

What has been done to enhance the interfacial properties of nanotubes?

Several methods have been used to modify nanotube surfaces chemically (Fig. 3, overleaf). One approach is to attach chemical groups covalently. Unfortunately, this strategy creates defects in the nanotube lattice, which can lower the electrical and thermal conductivity of the nanotubes and compromise their mechanical properties. Alternatively, non-covalent interactions can be used to either wrap surfactants or polymers around nanotubes, or to adsorb aromatic structures to the side-walls. These surface-modification methods are milder than the covalent approach, and usually cause little disruption to the structure or conductivity of the nanotubes. The idea of chemically modifying fillers to increase their interfacial strength isn't new — the surfaces of carbon fibres are also modified before use. But the near-perfect structures of nanotubes make the process more

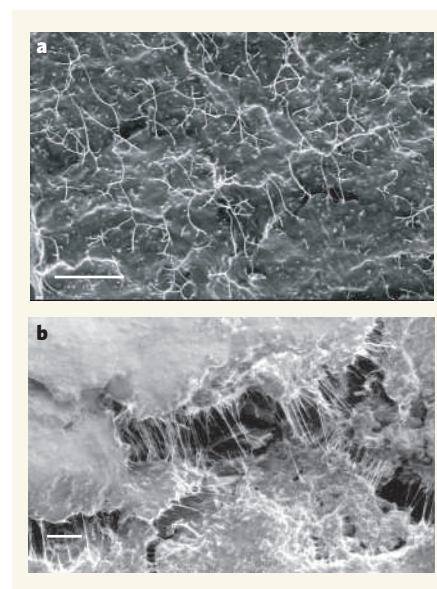


Figure 2 | Fractured composites. The reinforcing effect of nanotubes in composites can be undermined by weak interfacial contacts between nanotubes and the surrounding polymer matrix. **a**, In this scanning electron microscope (SEM) image of the fracture surface of a nanotube-epoxy composite, the fibres (seen in white contrast) are single-walled nanotubes that have been pulled out of the matrix and have then fallen back onto the fracture surface. **b**, This SEM image shows a crack in a nanotube-epoxy composite. Bundles of single-walled nanotubes (strands with pale contrast) are exposed and stretched between cracks in the composite. Scale bars, 1 µm.

challenging, as there are no defects upon which to anchor a pendent group, and nanotubes are chemically quite inert. Nevertheless, recent work in this area shows that nanotube-surface modification is cleaner and more controllable than carbon-fibre modification.

Couldn't the interfacial slippage be exploited?

Although interfacial shear is detrimental to stiffness and strength, it could produce high mechanical damping, as the large number of nanotube-polymer interfaces involved in nanotube slippage creates a lot of friction and energy dissipation. Viscoelastic behaviour at the interfaces causes this effect. Such damping is vital for many commercial applications, such as isolating objects from vibration. Once again, good dispersion of nanotubes in the matrices is essential to optimize this effect.

Is it possible to order the arrangement of nanotubes in composites?

Certainly. Nanotubes have been aligned in composites using magnetic fields, but this is unlikely to be especially useful because of the very high field strengths required. The best way to orient nanotubes is to put them under shear, for example by extrusion. Moving from randomly aligned mixtures of nanotubes in composites to well-ordered arrays can enhance the resulting material's performance under stress, and improve its electrical and thermal conductivities (conductivity in nanotubes is directional, being much better down their long axes). Such alignment also minimizes the number of nanotubes required to exert their effects. Perhaps most impressively, it is possible to incorporate aligned arrays of nanotubes into 'three-dimensional' composites. Traditional fibre-reinforced composites consist of stacked panels in which the fibres lie in plane; these materials have excellent in-plane properties, but fare poorly along the through-thickness direction. Inter-laminar nanotube arrays, aligned perpendicular to the stack panels, provide enhanced properties along the thickness direction. The arrays do this by fastening adjacent fibre stacks together — acting almost like Velcro between them.

We've seen that nanotube composites combine strength and conductivity — what else can they do?

With their wide range of excellent properties, nanotubes could act as fillers in diverse multifunctional composites. For example, the strength and damping properties of these materials might be increased if the nanotube-polymer interfaces can be suitably engineered. Adding nanotubes to plastics provides materials that could be used to shield microelectronics from electromagnetic interference. Similarly, experiments show that the microwave-absorbing capability of nanotubes could be exploited to heat temporary housing structures, and may have applications in space

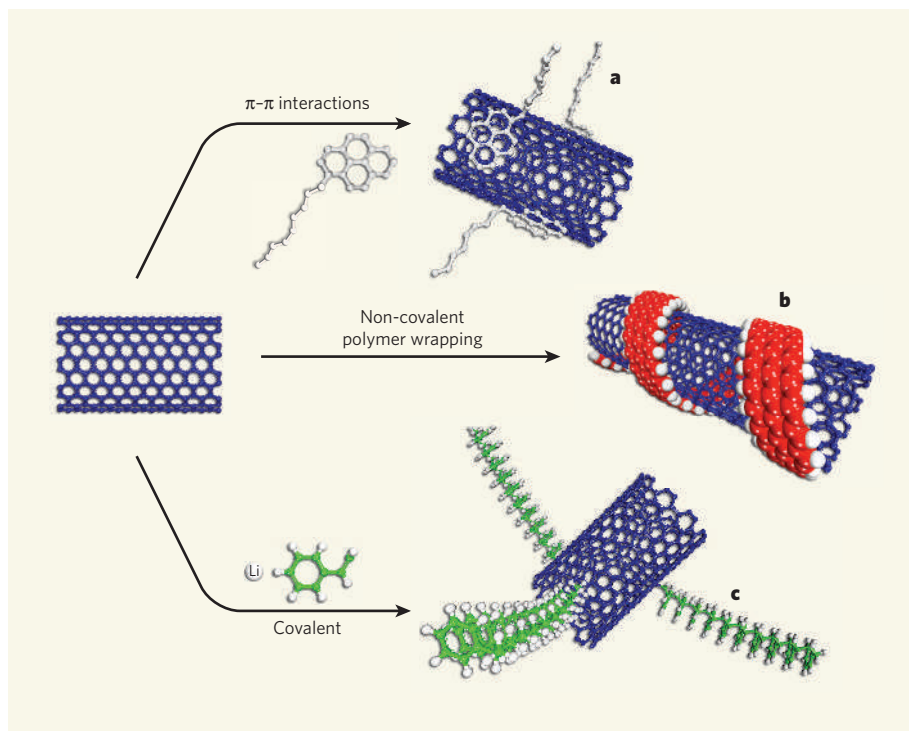


Figure 3 | Chemical modification of carbon nanotubes. Carbon nanotubes have poor solubility in most solvents and polymers, but this problem can be overcome by chemically modifying their surfaces. At present, three methods are used for this purpose. **a**, Aromatic molecules can be appended to nanotubes using certain non-covalent interactions (known as π - π interactions). Groups emanating from these molecules interact with the surrounding solvent or matrix. **b**, Non-covalent interactions (including π - π interactions, van der Waals forces and charge-transfer interactions) can be used to wrap polymers around nanotubes. **c**, Chemical groups can be covalently attached to nanotubes. In this case, lithium (Li) forms anions on the nanotube, which react with styrene monomers to form polystyrene chains covalently attached to the nanotube walls.

exploration. Furthermore, applying mechanical stress to nanotubes modulates their electrical conductivity. This effect could be exploited in advanced materials that have embedded sensors and actuators made of nanotubes, which could be used for real-time monitoring of stress distribution and to actively control the composite structures. Finally, thin nanotube layers might also be used in transparent conducting composites.

How will we see nanotube composites used in the future?

At present, the most widespread use is in electrostatic-discharge components, in which multiple-walled nanotubes slightly enhance the electrical conductivity of plastics. Some manufacturers of high-end sporting goods, such as tennis rackets, also claim to make their products from nanotube composites, so delivering superior performance. In the short term, the biggest markets for nanotube composites will undoubtedly be for high-value applications that can absorb the added costs — sectors such as aerospace (which needs lightweight, high-strength, high-temperature-resistant composites) and energy (for example, in nanotube-reinforced rubber seals for large oil-recovery platforms). Aerospace composites that have the required properties are already being developed. Once the cost of nanotubes

becomes comparable to that of carbon fibres (or even to that of the much cheaper reinforcer, carbon black), commodities such as nanotube-filled rubber tyres could become a reality. These are exciting times for nanotube composites. With strategies falling into place to solve the problems of their manufacture, commercial success for these materials — at least in select, value-added applications — is finally on the horizon.

Pulickel M. Ajayan is in the Department of Materials Science and Engineering, Rensselaer Polytechnic Institute, Troy, New York 12180, USA. James M. Tour is in the Departments of Chemistry and of Mechanical Engineering and Materials Science, and the Smalley Institute for Nanoscale Science and Technology, Rice University, MS 222, 6100 Main Street, Houston, Texas 77005, USA.
e-mails: ajayan@rpi.edu; tour@rice.edu

FURTHER READING

- Ajayan, P. M., Schadler, L. S., Giannaris, C. & Rubio, A. *Adv. Mater.* **12**, 750–754 (2000).
- Harris, P. J. F. *Int. Mater. Rev.* **49**, 31–43 (2004).
- Breuer, O. & Sundararaj, U. *Polym. Composites* **25**, 630–645 (2004).
- Dyke, C. A. & Tour, J. M. *J. Phys. Chem. A* **108**, 11151–11159 (2004).
- Yakobson, B. I. & Couchman, L. S. *J. Nanoparticle Res.* **8**, 105–110 (2006).
- Tasis, D., Tagmatarchis, N., Bianco, A. & Prato, M. *Chem. Rev.* **106**, 1105–1136 (2006).

Isotopic portrayal of the Earth's upper mantle flow field

Christine M. Meyzen¹, Janne Blichert-Toft¹, John N. Ludden², Eric Humler³, Catherine Mével⁴ & Francis Albarède¹

It is now well established that oceanic plates sink into the lower mantle at subduction zones, but the reverse process of replacing lost upper-mantle material is not well constrained. Even whether the return flow is strongly localized as narrow upwellings or more broadly distributed remains uncertain. Here we show that the distribution of long-lived radiogenic isotopes along the world's mid-ocean ridges can be used to map geochemical domains, which reflect contrasting refilling modes of the upper mantle. New hafnium isotopic data along the Southwest Indian Ridge delineate a sharp transition between an Indian province with a strong lower-mantle isotopic flavour and a South Atlantic province contaminated by advection of upper-mantle material beneath the lithospheric roots of the Archaean African craton. The upper mantle of both domains appears to be refilled through the seismically defined anomaly underlying South Africa and the Afar plume. Because of the viscous drag exerted by the continental keels, refilling of the upper mantle in the Atlantic and Indian domains appears to be slow and confined to localized upwellings. By contrast, in the unencumbered Pacific domain, upwellings seem comparatively much wider and more rapid.

Radiogenic isotope studies of Sr, Nd and Pb in oceanic islands and mid-ocean ridge basalts (MORB) have established the existence of distinct broad mantle isotopic provinces, such as the archetypal DUPAL anomaly^{1–3}. The distinctive isotopic properties of these large-scale domains are particularly conspicuous in the Sr, Nd, Pb and Hf isotope compositional space of MORB, as shown in Fig. 1. Remarkably, the mantle sources for Indian, South Atlantic, and Pacific MORB are distinct from that of North Atlantic MORB in having lower ε_{Hf} values. Compared to Atlantic and Indian MORB, Pacific MORB on average have lower $^{87}\text{Sr}/^{86}\text{Sr}$ values at a given ε_{Hf} . Indian MORB are characterized by higher $^{207}\text{Pb}/^{206}\text{Pb}$ and $^{208}\text{Pb}/^{206}\text{Pb}$ than their Atlantic and Pacific counterparts. The distinct isotopic identities of each province reflect the long-term segregation of mantle domains composed of various proportions of primordial mantle, melting residues, and recycled components of different ages and origins. Obtaining a clear picture of the transitional regions separating these provinces should help us to understand their genesis and role in global mantle convection patterns. The eastern boundary of the Indian province, which separates the Indian Ocean from the Pacific Ocean mantle domains, has been by far the most closely studied isotopic discontinuity. This boundary, which is located along the Australian–Antarctic Discordance (AAD), is extremely sharp, with $^{206}\text{Pb}/^{204}\text{Pb}$, $^{208}\text{Pb}/^{204}\text{Pb}$, and $^{176}\text{Hf}/^{177}\text{Hf}$ changing dramatically from Indian to Pacific values over less than 25 km on axis^{4,5} (Fig. 2). For this reason, this feature has been interpreted as reflecting a ‘convection front’ overlying either a mantle downwelling⁴, the upwelling of a cold but neutrally buoyant subducted slab associated with a shallow seismic anomaly⁶, or the presence of continental crust remnants⁵.

Because the isotope systematics of MORB from the AAD have proved to be informative about upper-mantle dynamics, we focus here on the isotope characteristics of MORB from the more gradual transition^{7,8} between the Indian and Atlantic domains along the ultraslow Southwest Indian Ridge (SWIR). The occurrence of South-Atlantic-like MORB is mainly restricted to the west of 26° E on this ridge, while Indian-like MORB are present in and to the east

of the Andrew Bain fracture zone ($\sim 30^\circ \text{E}$)^{7,8}. This transition between Indian-like and Atlantic-like mantle sources is broadly centred on the Andrew Bain fracture zone ($\sim 30^\circ \text{E}$). This fracture zone has been proposed to be associated with a diffuse boundary (26.31–32.89° E) separating the African plate into the Nubian and Somalian plates⁹ and we suggest that the distribution of radiogenic tracers along the SWIR is consistent with this view. A previous interpretation of the southward-decreasing $^{176}\text{Hf}/^{177}\text{Hf}$ gradient observed along the Mid-Atlantic Ridge as a consequence of the northwards spilling of mantle material from the Indian Ocean into the Atlantic mantle¹⁰ seems less efficient at explaining this distribution.

Hf and Nd isotope compositions in oceanic basalts are usually well correlated. Their occasional decoupling with respect to the average mantle array, such as along the AAD⁵ or the Mid-Atlantic Ridge north of Iceland¹¹, is symptomatic of unusual fractionation conditions during melting (for example, residual garnet) and thus provides a unique perspective on the dynamics of mantle isotopic provinces. Here, we report new Hf (and some Nd) isotopic compositions for about sixty samples collected between 35° (north of Marion Island) and the Rodrigues triple junction along the neovolcanic zone of the SWIR. These data supplement existing major-element and trace-element¹² and Sr, Nd, and Pb isotope compositions¹³ for the same samples. Here we interpret our new data in the context of global isotopic variability along the mid-ocean ridge system, in particular separating the geochemical signal of hotspots from that of the upper-mantle background. The strongly regionalized character of the asthenospheric isotopic signal is then discussed within the reference frame of the seismic-wave structure of the mantle beneath Africa and is shown to be consistent with the upper-mantle flow field as inferred by ref. 14 from shear-wave splitting and other geophysical observations.

The isotopic transition between Indian and South-Atlantic ridges

To provide a global framework for the examination of the present SWIR data, they were plotted together with Pb, Sr, Nd and Hf isotope data from the PetDB database (www.petdb.org) and other sources

¹Laboratoire des Sciences de la Terre, CNRS UMR 5570, Ecole Normale Supérieure de Lyon, Université Claude Bernard Lyon 1, 46 Allée d'Italie, 69364 Lyon cedex 07, France. ²British Geological Survey, Keyworth, NG12 5GG, Nottingham, UK. ³Laboratoire de Planétologie et Géodynamique, CNRS UMR 6112, Université de Nantes, 2 rue de la Houssinière, B.P. 92208, 44322 Nantes cedex 03, France. ⁴Laboratoire de Géosciences Marines, CNRS UMR 7154, Institut de Physique du Globe, 4 place Jussieu, 75252 Paris cedex 05, France.

(see Supplementary Information) covering the mid-ocean ridge system from the North Atlantic to the Juan de Fuca ridges (Fig. 2). Alternative pathways were possible, such as including the central Indian ridge, but because our primary goal was to obtain a first-order image of the global isotopic variability, we chose to focus on those ridge sections with the densest sampling coverage. This representation depicts, for the first time, the one-dimensional isotopic variability along most of the terrestrial mid-ocean ridge accretionary system. In Fig. 2, the angular distance along the axis was calculated using the successive Euler poles of opening of the consecutive seafloor segments with respect to an arbitrary reference point in the North Atlantic (77.53° N, 7.67° E, see Supplementary Information for explanation of the relevant calculation). The reported isotopic data were first reduced by subtracting the mean value for the entire ridge system and dividing the result by the standard deviation. A new parameter based on Pb isotopes that is particularly suitable for detecting the Indian isotopic affinity of any given sample is also introduced in Fig. 2 with the dual justification that (1) the DUPAL character of the Indian Ocean province is most visible in terms of Pb isotopes¹ (Fig. 1d), and (2) Pb isotope compositions in MORB all fan out from C in Pb isotope space¹⁵ (Fig. 1d). This new parameter R_C was defined as:

$$R_C = \frac{(^{208}\text{Pb}/^{206}\text{Pb}) - (^{208}\text{Pb}/^{206}\text{Pb})_C}{(^{207}\text{Pb}/^{206}\text{Pb}) - (^{207}\text{Pb}/^{206}\text{Pb})_C}$$

in which C refers to the Pb common composition of ref. 15, which is essentially equivalent to the FOZO (Focus Zone) component assigned by ref. 16 to the lower mantle (Fig. 1d). The parameter R_C reflects the time-integrated parent–daughter Th/U ratio in the mantle source of oceanic basalts in much the same way as, for example, $^{143}\text{Nd}/^{144}\text{Nd}$ reflects the Sm/Nd ratio and $^{206}\text{Pb}/^{204}\text{Pb}$ reflects the U/Pb ratio. It thus captures the higher values of $^{208}\text{Pb}/^{206}\text{Pb}$ for a given $^{207}\text{Pb}/^{206}\text{Pb}$,

and hence the higher time-integrated Th/U of Indian Ocean MORB with respect to those of other MORB (Figs 1d and 2).

The diagrams in Fig. 2 have been divided into four main regions: the North and central Atlantic (I), the South Atlantic (II), the southwest and southeast Indian (III), and the Pacific (IV) domains. Regions I and IV will not be discussed in detail here and may admittedly be subdivided further, such as in region I south of the Azores. To assist the discussion, the $^{206}\text{Pb}/^{204}\text{Pb}$ of the C component¹⁵ and the R_C value of the depleted mantle component¹³ are also shown in Fig. 2.

The unradiogenic end of the north–south decreasing gradient of $^{176}\text{Hf}/^{177}\text{Hf}$ defined by Atlantic MORB^{7,10,17} separates the South-Atlantic-like (II) from the southwest-Indian-Ocean-like (III) mantle domains along the SWIR at an angular distance of $\sim 150^\circ$ with respect to the reference point defined above (Fig. 2). This boundary coincides with abrupt changes in R_C , $^{206}\text{Pb}/^{204}\text{Pb}$, and $^{87}\text{Sr}/^{86}\text{Sr}$ and manifests itself in the $^{143}\text{Nd}/^{144}\text{Nd}$ profile as a steep gradient opposite to the gradient in the South Atlantic. The main reason for isolating region II in the South Atlantic from region I in the North Atlantic is the presence of breaks observed north of Tristan da Cunha at $\sim 20^\circ$ S (ref. 18) in the $^{143}\text{Nd}/^{144}\text{Nd}$, $^{87}\text{Sr}/^{86}\text{Sr}$ and R_C profiles.

Unscrambling the hotspot and asthenospheric signals

Interaction of mid-ocean ridges with hotspots has been identified in a number of different localities, such as near Bouvet Island on the southernmost Mid-Atlantic Ridge, the America–Antarctic ridge, and the westernmost SWIR¹⁹, which all exhibit extreme geochemical variability. As will be discussed below, this variability is independent of the nature of the geochemical components present in MORB and in plume magmas. To highlight the effect of hotspots on the isotopic composition of MORB, the standard deviation of each isotopic parameter j (for example, $j = ^{143}\text{Nd}/^{144}\text{Nd}$) was calculated over 5°

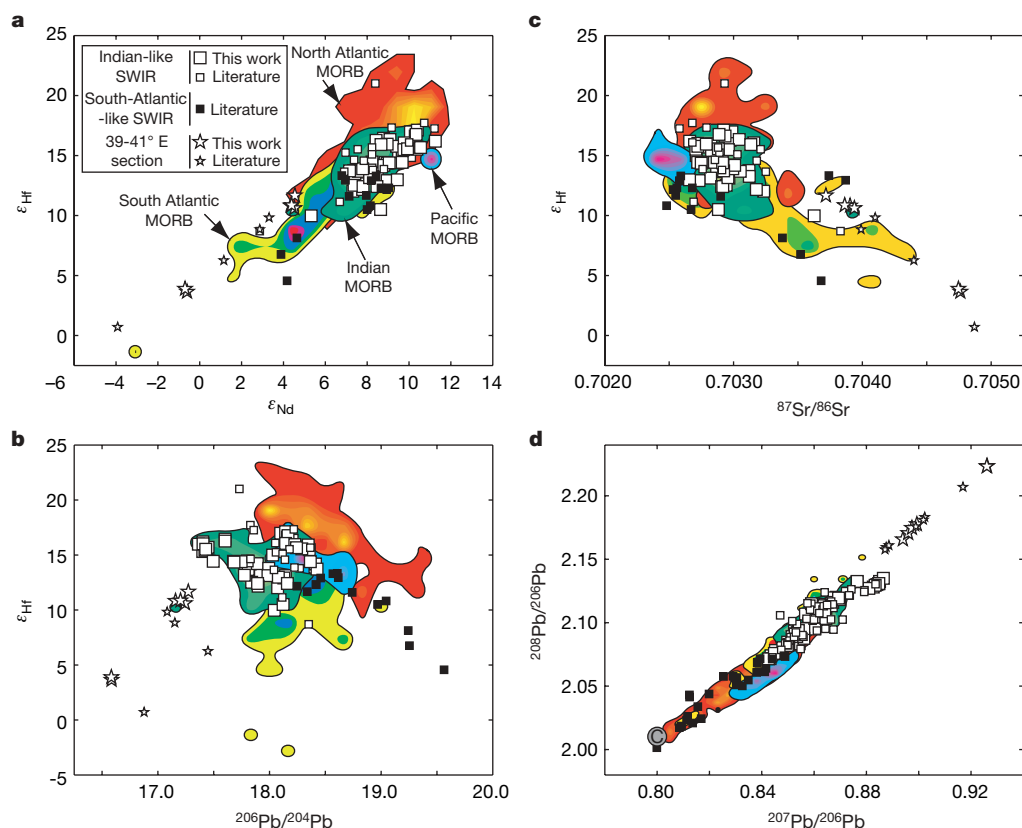


Figure 1 | Isotopic characteristics of SWIR lavas compared to those of other MORB from the Indian, Pacific, South Atlantic and North Atlantic oceans. a, ϵ_{Hf} versus ϵ_{Nd} . b, ϵ_{Hf} versus $^{206}\text{Pb}/^{204}\text{Pb}$. c, ϵ_{Hf} versus $^{87}\text{Sr}/^{86}\text{Sr}$. d, $^{208}\text{Pb}/^{206}\text{Pb}$ versus $^{207}\text{Pb}/^{206}\text{Pb}$. SWIR data for Pb, Sr and Nd isotopes are

from ref. 13. Colour-coded fields are two-dimensional histograms calculated from MORB data retrieved from the PetDB resource and from recent literature (see the Supplementary Information for data sources) as well as the data of this work. C denotes the common component of ref. 15.

intervals i (s_i^j , where $i = 1, 2, \dots$) as a function of the angular distance along the ridge axis and normalized to the total standard \bar{s}_j deviation of the whole MORB population considered in the compilation. Such a representation gives a dimensionless estimate of the local isotopic variability for each parameter (Fig. 3a). The observed isotopic variability is concentrated around a small number of peaks broadly coinciding with the hotspots of the islands of Iceland, the Azores, and the Bouvet, St Paul and Salas y Gomez. With the exception of the ridge section between Tristan da Cunha and Bouvet Island, the background variability is minimal between peaks regularly spaced at $30\text{--}40^\circ$. Spectral analysis is an alternative tool to estimate the hotspot spacing along mid-ocean ridges. The parameter j was first reduced by subtracting the mean value and normalizing to \bar{s}_j . The reduced signals obtained in this way are dimensionless and can therefore be added ('stacked'). The spectra of the reduced signals were then calculated using Lomb–Scargle periodograms, which are suitable for

very irregularly spaced samples²⁰. Because the peak positions look rather similar for each parameter j (see Supplementary Information), only the spectrum of the stacked reduced signals is shown in Fig. 3b.

The three strongest peaks at $1/0.004^\circ$, $1/0.023^\circ$ and $1/0.032^\circ$ correspond to dominant wavelengths at 250° , 43° and 31° , respectively. The signal at 250° reflects the overall larger hotspot abundance in the Atlantic and Indian Oceans relative to the Pacific Ocean. The wave-numbers of the other two peaks are shifted by approximately $(0.032 - 0.023)/2 = 1/0.0045^\circ$, which suggests that the 43° and 31° peaks represent a modulation of a single wavelength at $\sim 35^\circ$ by the 250° signal. The origin of this periodicity is unknown but may reflect the dominance of some particular convective modes (which can be defined as the number of maxima or minima of the upwelling velocity on a great circle) along the mid-ocean ridge system. The signal is, however, strong enough on the isotopic profiles to separate the hotspot signal from its asthenospheric background dominated by

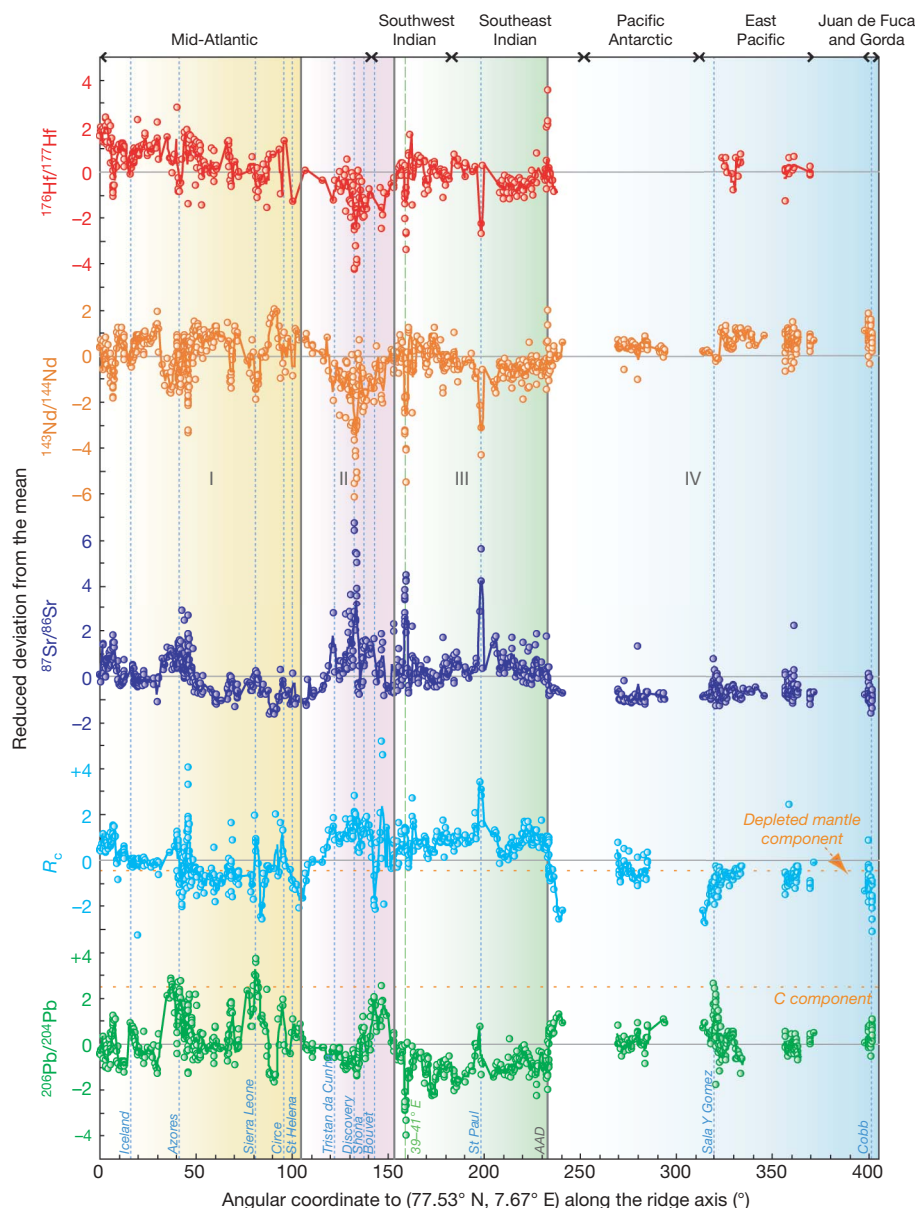


Figure 2 | Variations of $^{176}\text{Hf}/^{177}\text{Hf}$, $^{143}\text{Nd}/^{144}\text{Nd}$, $^{87}\text{Sr}/^{86}\text{Sr}$, R_c and $^{206}\text{Pb}/^{204}\text{Pb}$ along the ridge axis from the North Atlantic to Juan de Fuca. Locations of ridges and major hotspots are also shown, including those of the $39\text{--}41^\circ\text{E}$ section and the AAD. See the Supplementary Information for data description and data sources. The northernmost sample in the Atlantic Ocean (77.53°N , 7.67°E) is the global reference point. Four regions have

been identified: the North and central Atlantic (I), the South Atlantic (II), the Southwest and Southeast Indian (III), and the Pacific (IV). The curve corresponds to a smoothing of the data using a gaussian filter with a 0.5° standard deviation. Points with uncertainties in excess of 20% on R_c have been left out.

geochemically depleted material. Peaks of strong isotopic variability will thus be assigned to hotspots, while the baseline will be considered as representing the MORB source.

Geochemical MORB components

The origin of the geochemical components that define the South Atlantic and Indian 'flavours' of domains II and III, respectively, is not fully understood. The high $^{206}\text{Pb}/^{204}\text{Pb}$ 'HIMU' component, typical of just a small number of islands, such as St Helena, is restricted to the neighbourhood of a few hotspots, such as Shona²¹, and does not seem to be part of the MORB background. The low $^{206}\text{Pb}/^{204}\text{Pb}$ (low $^{238}\text{U}/^{204}\text{Pb}$ or 'LOMU') component has attracted

considerable attention and four interpretations have been put forward, each corresponding to the presence of various components in the MORB source: (1) subducted oceanic crust and sediments^{22–24} and dehydration fluids^{25,26}, (2) debris of subcontinental lithosphere trailing in the asthenosphere after the dislocation of the Gondwana supercontinent over the past 200 million years^{8,21,27–29}, (3) debris of delaminated lower crust^{5,13,30,31} or (4) input from the lower mantle^{2,29}. The presence of subcontinental lithosphere modified by subduction—that is, a combination of interpretations (1) and (2)—was also considered a possibility for the SWIR⁷. A record of ancient subduction processes, dragging down debris from both the oceanic crust and the lower continental crust now stored in the lithospheric keels of Archaean cratons, is provided by the occurrence of eclogites among xenoliths from South African kimberlites³².

Between Bouvet Island and Tristan da Cunha (region II in Fig. 2), the South Atlantic Ridge is underlain by asthenosphere with high $^{87}\text{Sr}/^{86}\text{Sr}$, low $^{143}\text{Nd}/^{144}\text{Nd}$ and $^{176}\text{Hf}/^{177}\text{Hf}$, and high R_C values, suggesting that it has incorporated continental crust debris. However, a strong gradient in Hf isotopes as observed on this segment by refs 10 and 17 indicates that the contribution of continental crust decreases towards the central Atlantic. Likewise, an extreme isotopic variability within the Indian domain, notably between 39 and 41° E of longitude¹³ and on the Indian side of the AAD⁵, hints at a local presence of lower crustal debris trailing in the mantle following the Gondwana break-up.

Apart from these conspicuous but local anomalies, the isotopic properties of the Indian-type MORB upper mantle are inconsistent with the ubiquitous addition of common compositions of a crustal component⁷: Hf is not less radiogenic in Indian MORB than in Pacific MORB and ϵ_{Hf} does not correlate with $^{87}\text{Sr}/^{86}\text{Sr}$ (Fig. 1). Furthermore, the presence of subducted sediments²⁴ in the source of Indian MORB is incompatible with the unradiogenic Pb^{31,33} of the latter. Rather, the unmistakable common isotopic flavour of ocean island and mid-ocean ridge basalts from the Indian Ocean domain (region III) suggests a deep mantle origin for the overall Indian signature^{2,29} (see Supplementary Information). As is suggested by the extreme geochemical variability of ocean island basalts¹⁸, the lower mantle is certainly not homogeneous. Remarkably, the isotopic compositions of Nd, Sr, Hf and Pb of basalts from the Hawaiian hotspot^{34,35} plot either within the field of Indian MORB of Fig. 1 or at its enriched end and this is also true for some other hotspots. A comparable connection is borne out by the high R_C values of most Indian MORB, which reflect high time-integrated parent–daughter Th/U ratios exactly as in the source of ocean island basalts³⁶ (Fig. 2). A hotspot-like component therefore seems to be present in the Indian Ocean asthenosphere.

Upper-mantle flow field from the Indian and South Atlantic oceans

The mechanisms that led to the establishment in the upper mantle of a huge isotopic province with a lower-mantle signature and occupying most of the Indian upper mantle can be explained by simple geodynamic constraints. It is unlikely that the Indian isotopic province owes its existence to very-low-order modes of mantle convection. The broad latitudinal band of mantle with isotopically similar properties such as hypothesized by ref. 1 is inconsistent with the steep juxtaposition of three isotopic provinces (regions II, III, and IV) at similar latitudes. Furthermore, an order-3 mode (as inferred from ref. 2) in the Southern Hemisphere is incompatible with the strong contrast between both Indian and South Atlantic MORB and Indian and South Pacific MORB and with the sharp transition between these isotopic provinces.

Instead, the provinciality of the Indian Ocean (region III), coupled with its isotopic characteristics, is more consistent with replenishment of the upper mantle through a localized upwelling, a mechanism which is also in agreement with the seismic structure of the underlying mantle. An inclined low-velocity anomaly extends from the core–mantle boundary region beneath the southeastern Atlantic

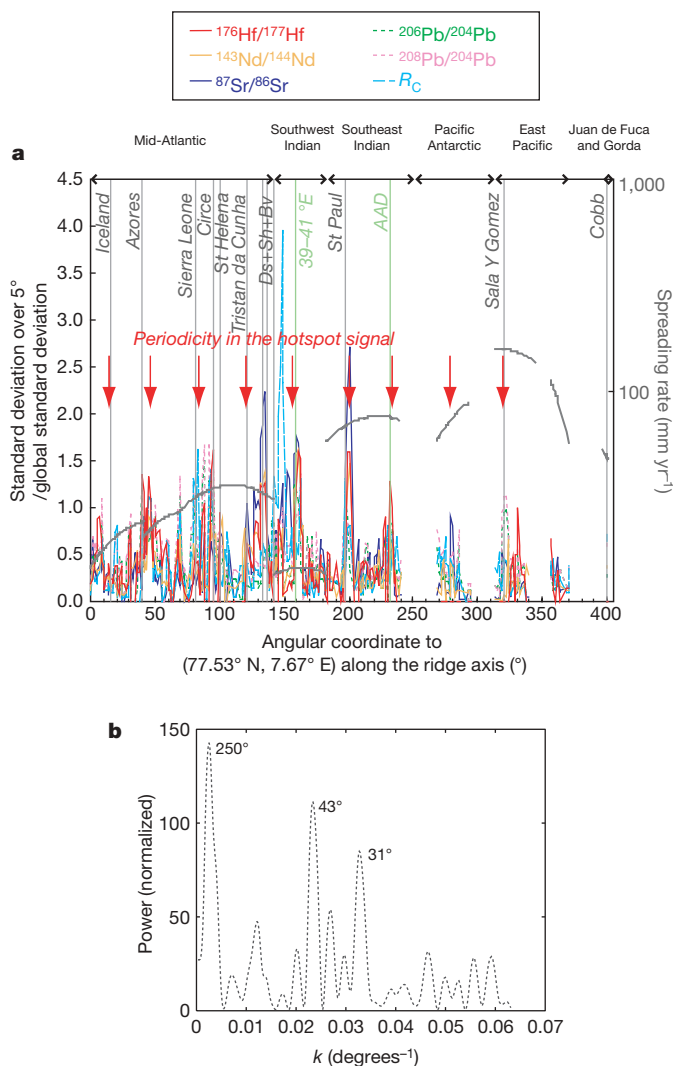


Figure 3 | Spectra of the isotopic signals. **a**, Variations along the ridge axis of the standard deviation of isotopic ratios calculated over 5° intervals and normalized to the global value. Locations of hotspots and mid-ocean ridges are also shown. Ds+Sh+Bv indicates the Discovery, Shona and Bouvet hotspots, respectively. Spreading rates (grey curves) are from ref. 52. This plot shows that (1) with the exception of the segment between Bouvet Island and Tristan da Cunha, the variance peaks are nearly regularly distributed (at 35–40° spacing, red arrows), and (2) there is little or no correlation between isotopic variance and spreading rate. **b**, Plot of the periodograms of the stacked isotopic signals with respect to the wavenumber k in units of inverse degrees. The peak at 250° emphasizes the global contrast between the Pacific and the Atlantic + Indian domains. The two peaks at 43° and 31° represent a modulation of a single wavelength at ~35° by the 250° signal. The 35° signal reflects that the isotopic effect of ridge–hotspot interaction is regularly distributed along the ridge axis. See the Supplementary Information for data sources.

Ocean into the upper mantle beneath East Africa^{37–39}. Using mantle velocity inferences from shear-wave splitting analysis, ref. 14 defined the three-dimensional flow field created by this anomaly and found it to be dominated by a major upwelling originating in the lower mantle beneath southern Africa and radiating along the base of the asthenosphere under the Afar plume. Strong independent support for this model is provided by the presence of an Indian isotopic signature in Gulf of Aden basalts, which are strongly influenced by the Afar plume, at least west of 48° E (ref. 40).

Both geochemical and geophysical evidence therefore confirm that lower-mantle material from the African upwelling flows into the asthenosphere underlying East Africa and the Indian Ocean. The contaminated mixture then spreads southwards along the Indian ridges (Fig. 4) and only reaches the South Atlantic Ridge after first having been deflected under the deep roots of the South African Archaean cratons, extending down to depths in excess of 300 km (refs 41, 42). Continental roots contaminated by old subduction zones or sprinkled with subducted fragments and continental crust³² impart a distinct continental signature on the mantle source of South Atlantic MORB. A similar mechanism has recently been proposed to account for the anomalous isotopic compositions of the Walvis ridge⁴³. The relatively low ³He/⁴He values⁴⁴ observed along the South Atlantic Ridge are consistent with a He contribution from the subcontinental lithospheric mantle, which is known to be more radiogenic than asthenosphere He^{44,45}. Likewise, on the SWIR at 39–41° E, the continental signature of this segment reflects the effect of a southward mantle flow grazing the roots of the Madagascar craton.

The ‘holes-in-the-floor’ model¹⁴, describing exchange of material across the transition zone, is consistent with available geochemical, geodynamic, and geophysical evidence. Upper-mantle material lost to the lower mantle, mostly as rigid plates dragged down by their own weight, is replaced by upwelling of lower-mantle material through ‘holes’ of different dimensions in the 660 km discontinuity, such as that underlying East Africa¹⁴. Ref. 46 argued that mantle does not flow freely across the transition zone, which behaves as a leaky boundary with respect to mantle convection with upwellings penetrating more easily than downwellings⁴⁷. This view is supported by the assessment of mid-mantle heat flow based on seismic tomography, which suggests that plumes account for all upward advective heat transport in the lower mantle that eventually breaks through into the upper mantle⁴⁸.

In addition, the Gondwana supercontinent and its break-up history seem to have played a major role in the building of the

present-day asthenospheric flow field pattern. In ref. 49 the Pacific (Panthalassan, region IV in Fig. 2) domain, which contains only oceanic plates, is opposed to the Atlantic–Indian (Pangean, regions I to III) domain, which contains all the continental plates. Compared to the Pacific domain, the mid-ocean ridges from the South Atlantic and Indian Oceans are characterized by (1) abundant near-ridge hotspots, such as Tristan da Cunha and the Bouvet and St Paul islands, correlated with strong peaks of isotopic variability (Fig. 3a), and (2) slow spreading velocities (Fig. 3a) and therefore weaker heat loss. As argued in ref. 14, the upper mantle in the Gondwana domain is contaminated by multiple upwellings, and the drag exerted by the deep roots of the Archaean cratons strongly perturbs the velocity field. This mechanism is also expressed around the ultraslow SWIR, where the thick keels of the African⁴² and Antarctic cratons resist the asthenospheric flow.

In contrast, the weak isotopic variability and the fast spreading rates observed in the Pacific ridge system (region IV) are best interpreted by the presence of broad upper-mantle upwellings with no hindrance from continental lithosphere. Such a pattern is inconsistent with the focusing of broad upwellings under large expanses of continental lithosphere inferred from their insulating effect on mantle convection^{50,51}. In contrast, the viscous drag exerted by the continental keels seems to oppose the rapid rifting of continents. Erosion of continental keels is, at least under southwestern Africa⁴¹, a slow process. A strong viscous drag also explains why heat builds up in the deep mantle under the South Atlantic Ocean and its neighbouring regions, thereby giving this part of the mantle anomalous seismic properties³⁹ and promoting the generation of a large number of small hotspots.

Received 4 March; accepted 10 May 2007.

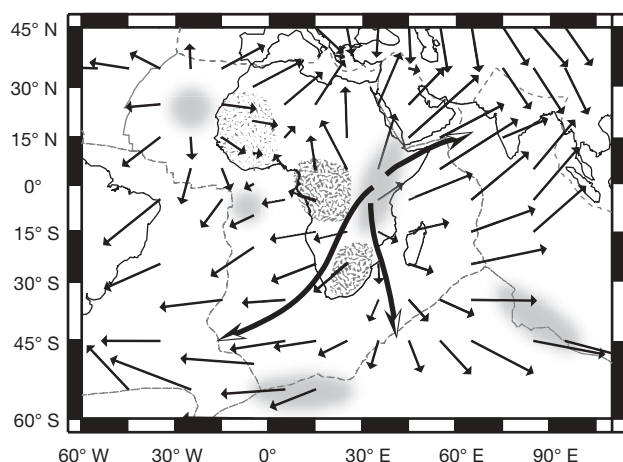


Figure 4 | Map view of the plate- and density-driven flow field at the base of the asthenosphere adapted from figure 7 of ref. 14 with permission. The base of the asthenosphere is at a depth of 300 km. Arrows illustrate horizontal flow. The shaded fields indicate upwellings. Note the radial flow field at the base of the asthenosphere as underlined by bold thick arrows. The patterned areas show the positions of the African cratons.

- Hart, S. R. A large-scale isotope anomaly in the Southern Hemisphere mantle. *Nature* **309**, 753–757 (1984).
- Castillo, P. The DUPAL anomaly as a trace of the upwelling lower mantle. *Nature* **336**, 667–670 (1988).
- Dupré, B. & Allègre, C. Pb–Sr isotope variation in Indian Ocean basalts and mixing phenomena. *Nature* **303**, 142–146 (1983).
- Klein, E. M., Langmuir, C. H., Zindler, A., Staudigel, H. & Hamelin, B. Isotope evidence of mantle convection boundary at the Australian–Antarctic Discordance. *Nature* **333**, 623–629 (1988).
- Hanan, B. B., Blichert-Toft, J., Pyle, D. G. & Christie, D. M. Contrasting origins of the upper mantle revealed by hafnium and lead isotopes from the Southeast Indian Ridge. *Nature* **432**, 91–94 (2004).
- Gurnis, M., Muller, R. D. & Moresi, L. Cretaceous vertical motion of Australia and the Australian–Antarctic Discordance. *Science* **279**, 1499–1504 (1998).
- Janney, P. E., Le Roex, A. P. & Carlson, R. W. Hafnium isotope and trace element constraints on the nature of mantle heterogeneity beneath the Central Southwest Indian Ridge (13°E to 47°E). *J. Petrol.* **46**, 2427–2464 (2005).
- Mahoney, J., Le Roex, A. P., Peng, Z., Fisher, R. L. & Natland, J. H. Southwestern limits of Indian Ocean ridge mantle and origin of low ²⁰⁶Pb/²⁰⁴Pb mid-ocean ridge basalt: Isotope systematics of the Central Southwest Indian Ridge (17–50°E). *J. Geophys. Res.* **97**, 19771–19790 (1992).
- Horner-Johnson, B. C., Gordon, R. G., Cowles, S. M. & Argus, D. F. The angular velocity of Nubia relative to Somalia and the location of the Nubia–Somalia–Antarctica triple junction. *Geophys. J. Int.* **162**, 221–238 (2005).
- Andres, M., Blichert-Toft, J. & Schilling, J.-G. Nature of the depleted upper mantle beneath the Atlantic: evidence from Hf isotopes in normal mid-ocean ridge basalts from 79°N to 55°S. *Earth Planet. Sci. Lett.* **225**, 89–103 (2004).
- Blichert-Toft, J. *et al.* Geochemical segmentation of the Mid-Atlantic Ridge north of Iceland and ridge-hot spot interaction in the North Atlantic. *Geochem. Geophys. Geosyst.* **6**, Q01E19, doi:10.1029/2004GC000788 (2005).
- Meyzen, C. M., Toplis, M. J., Humler, E., Ludden, J. N. & Mével, C. A discontinuity in mantle composition beneath the Southwest Indian Ridge. *Nature* **421**, 731–733 (2003).
- Meyzen, C. M. *et al.* New insights into the origin and distribution of the DUPAL isotope anomaly in the Indian Ocean mantle from MORB of the Southwest Indian Ridge. *Geochem. Geophys. Geosyst.* **6**, Q11K11, doi:10.1029/2005GC000979 (2005).
- Behn, M. D., Conrad, C. P. & Silver, P. G. Detection of upper mantle flow associated with the African Superplume. *Earth Planet. Sci. Lett.* **224**, 259–274 (2004).
- Hanan, B. B. & Graham, D. W. Lead and helium isotope evidence from oceanic basalts for a common deep source of mantle plume. *Science* **272**, 991–995 (1996).
- Hart, S. R., Hauri, E. H., Oschmann, L. A. & Whitehead, J. A. Mantle plumes and entrainment—Isotopic evidence. *Science* **256**, 517–520 (1992).

17. Agraniér, A. *et al.* The spectra of isotopic heterogeneities along the Mid-Atlantic Ridge. *Earth Planet. Sci. Lett.* **238**, 96–109 (2005).
18. Hofmann, A. W. in *In The Mantle and Core* (ed. Carlson, R. W.) 61–101 (Treatise on Geochemistry, Holland and Turekian, Oxford, 2003).
19. Kurz, M. D., Le Roex, A. P. & Dick, H. J. B. Isotope geochemistry of the enriched mantle near the Bouvet triple junction. *Geochim. Cosmochim. Acta* **62**, 841–852 (1998).
20. Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. *Numerical Recipes: The Art of Scientific Computing* 569–576 (University Press, Cambridge, 1992).
21. Douglass, J., Schilling, J.-G. & Fontignie, D. Plume–ridge interactions of the Discovery and Shona mantle plumes with the southern Mid-Atlantic Ridge (40°S to 55°S). *J. Geophys. Res.* **104**, 2941–2962 (1999).
22. Hamelin, B., Dupré, B. & Allègre, C. J. Pb–Sr–Nd isotopic data of Indian–Ocean ridges—New evidence of large-scale mapping of mantle heterogeneities. *Earth Planet. Sci. Lett.* **76**, 288–298 (1986).
23. Le Roex, A. P., Dick, H. J. B. & Fisher, R. L. Petrology and geochemistry of MORB from 25° to 46°E along the Southwest Indian Ridge—Evidence for contrasting styles of mantle enrichment. *J. Petrol.* **30**, 947–986 (1989).
24. Rehkämpfer, M. & Hofmann, A. W. Recycled ocean crust and sediment in Indian Ocean MORB. *Earth Planet. Sci. Lett.* **147**, 93–106 (1997).
25. Le Roux, P. J. *et al.* Mantle heterogeneity beneath the Southern Mid-Atlantic Ridge: trace element evidence for contamination of ambient asthenospheric mantle. *Earth Planet. Sci. Lett.* **203**, 479–498 (2002).
26. Kempton, P. D. *et al.* Sr–Nd–Pb–Hf isotope results from ODP Leg 187: evidence for mantle dynamics of the Australian–Antarctic Discordance and origin of the Indian MORB source. *Geochem. Geophys. Geosyst.* **3**, doi:10.1029/2002GC000320 (2002).
27. Mahoney, J. J., Natland, J. H., White, W. M., Poreda, R. & Bloomer, S. H. Isotopic and geochemical provinces of the Indian Ocean spreading centers. *J. Geophys. Res.* **94**, 4033–4052 (1989).
28. Andres, M., Blichert-Toft, J. & Schilling, J.-G. Hafnium isotopes in basalts from the southern Mid-Atlantic Ridge from 40°S to 55°S: Discovery and Shona plume–ridge interactions and the role of recycled sediments. *Geochem. Geophys. Geosyst.* **3**, 8502, doi:10.1029/2002GC000324 (2002).
29. Storey, M. *et al.* Contamination of Indian Ocean asthenosphere by the Kerguelen Heard mantle plume. *Nature* **338**, 574–576 (1989).
30. Kamenetsky, V. S. *et al.* Remnants of Gondwanan continental lithosphere in oceanic upper mantle: evidence from the South Atlantic Ridge. *Geology* **29**, 243–246 (2001).
31. Escrig, S., Capmas, F., Dupré, B. & Allègre, C. J. Osmium isotopic constraints on the nature of the DUPAL anomaly from Indian mid-ocean-ridge basalts. *Nature* **431**, 59–63 (2004).
32. Carlson, R. W. *et al.* Continental growth, preservation and modification in southern Africa. *GSA Today* **10**, 1–7 (2000).
33. Zhang, S.-Q. *et al.* Evidence for a widespread Tethyan upper mantle with Indian-ocean-type isotopic characteristics. *J. Petrol.* **46**, 829–858 (2005).
34. Blichert-Toft, J., Frey, F. A. & Albarède, F. Hf isotope evidence for pelagic sediments in the source of Hawaiian basalts. *Science* **285**, 879–882 (1999).
35. Bryce, J. G., DePaolo, D. J. & Lassiter, J. C. Geochemical structure of the Hawaiian plume: Sr, Nd, and Os isotopes in the 2.8 km HSDP-2 section of Mauna Kea volcano. *Geochem. Geophys. Geosyst.* **6**, Q09G18, doi:10.1029/2004GC000809 (2005).
36. Gasperini, D. *et al.* Evidence from Sardinian basalt geochemistry for recycling of plume heads into the earth's mantle. *Nature* **408**, 701–704 (2000).
37. Grand, S. P., VanDerHilst, R. D. & Widiyantoro, S. Global seismic tomography: A snapshot of convection in the Earth. *GSA Today* **7**, 1–6 (1997).
38. Ritsema, J., Ni, S., Helmberger, D. V. & Crotwell, H. P. Evidence for strong shear velocity reductions and velocity gradients in the lower mantle beneath Africa. *Geophys. Res. Lett.* **25**, 4245–4248 (1998).
39. Helmberger, D. V., Ni, S., Wen, L. & Ritsema, J. Seismic evidence for Ultra Low Velocity Zones beneath Africa and Eastern Atlantic. *J. Geophys. Res.* **105**, 23865–23878 (2000).
40. Schilling, J. G., Kingsley, R. H., Hanan, B. H. & McCully, B. L. Nd–Sr–Pb isotopic variations along the Gulf of Aden: Evidence for Afar mantle plume–continental lithosphere interaction. *J. Geophys. Res.* **97**, 10927–10966 (1992).
41. Ritsema, J., Nyblade, A. A., Owens, T. J., Langston, C. A. & VanDecar, J. C. Upper mantle seismic velocity structure beneath Tanzania, East Africa: Implications for the stability of cratonic lithosphere. *J. Geophys. Res.* **103**, 21201–21213 (1998).
42. James, D. E., Fouch, M. J., VanDecar, J. C., Van Der Lee, S. & Group, K. S. Tectospheric structure beneath Southern Africa. *Geophys. Res. Lett.* **28**, 2485–2488 (2001).
43. Class, C. & Le Roex, A. P. Continental material in the shallow oceanic mantle - How does it get there? *Geology* **34**, 129–132 (2006).
44. Graham, D. W. in *Noble Gases in Geochemistry and Cosmochemistry* (eds Porcelli, D., Wieler, R. & Ballentine, C.) 247–318 (Mineralogical Society of America, Washington DC, 2002).
45. Gautheron, C. & Moreira, M. Helium signature of the subcontinental lithospheric mantle. *Earth Planet. Sci. Lett.* **199**, 39–47 (2002).
46. Tackley, P. J., Stevenson, D. J., Glatzmaier, G. A. & Schubert, G. Effects of multiple phase transitions in a 3-dimensional spherical model of convection in Earth's mantle. *J. Geophys. Res.* **99**, 15877–15901 (1994).
47. Tackley, P. J. On the penetration of an endothermic phase transition by upwellings and downwellings. *J. Geophys. Res.* **100**, 15477–15488 (1995).
48. Nolet, G., Karato, S.-I. & Montelli, R. Plume fluxes from seismic tomography. *Earth Planet. Sci. Lett.* **248**, 685–699 (2006).
49. Collins, W. J. Slab pull, mantle convection, and Pangaeian assembly and dispersal. *Earth Planet. Sci. Lett.* **205**, 225–237 (2003).
50. Zhong, S. & Gurnis, M. Dynamic feedback between a continent like raft and thermal convection. *J. Geophys. Res.* **98**, 12219–12232 (1993).
51. Gurnis, M. Large-scale mantle convection and the aggregation and dispersal of supercontinents. *Nature* **332**, 695–699 (1988).
52. Gordon, R. G. *Present Plate Motions and Plate Boundaries* (ed. Ahrens, T.) 66–87 (AGU, Washington, 1995).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank P. Telouk for assistance with the Nu Plasma HR, and B. Reynard, I. Daniel, P. Oger and F. Chambat for suggestions. We also thank M. Behn for providing the map used to draw Fig. 4, J. Ritsema for sharing unpublished seismic profiles, and L. Dosso and B. Hanan for allowing us to view their unpublished data from the Pacific–Antarctic Ridge. Financial support of the analytical work and publication costs were provided by CNRS-INSU (the programmes DyETI and SEDIT), while IPEV provided access to the Marion Dufresne II for sampling during the EDUL (summer 1997) and SWIFT (winter 2001) cruises.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.M.M. (christine.meyzen@ens-lyon.fr).

Probabilistic reasoning by neurons

Tianming Yang¹ & Michael N. Shadlen¹

Our brains allow us to reason about alternatives and to make choices that are likely to pay off. Often there is no one correct answer, but instead one that is favoured simply because it is more likely to lead to reward. A variety of probabilistic classification tasks probe the covert strategies that humans use to decide among alternatives based on evidence that bears only probabilistically on outcome. Here we show that rhesus monkeys can also achieve such reasoning. We have trained two monkeys to choose between a pair of coloured targets after viewing four shapes, shown sequentially, that governed the probability that one of the targets would furnish reward. Monkeys learned to combine probabilistic information from the shape combinations. Moreover, neurons in the parietal cortex reveal the addition and subtraction of probabilistic quantities that underlie decision-making on this task.

Decision-making is a complicated process that is often based on more than one source of evidence. The brain needs to combine these sources to maximize the chance of achieving a correct decision or to achieve another related goal. Recent advances in neuroscience are beginning to expose the neurobiological mechanisms that underlie simple decisions^{1–6}. It has been demonstrated that, when the outcome of a decision is an eye movement, a neural correlate of the evolving decision can be recorded in brain areas associated with high level motor planning and attention allocation^{7–11}. More specifically, neurons in the lateral intraparietal area (LIP) have been shown to accumulate sensory information provided by earlier visual cortex when a decision is being formed^{8,9,12,13}. The mechanism mimics statistical decision processes that accrue evidence sequentially in the form of a log likelihood ratio (logLR) that favours one outcome over another^{4,14,15}. Therefore, it has been hypothesized that a neuronal substrate of probability integration exists in area LIP¹⁶.

To test this hypothesis, we trained two monkeys to perform a probabilistic categorization task (Fig. 1a). The task was adapted from the well-known weather-prediction task^{17,18} used to study human learning and memory. In each trial the monkey viewed four highly discriminable shapes; these were sampled randomly (with replacement) from a set of ten possible shapes. The shapes were added successively to the display over four half-second epochs. The monkey then made an eye movement to either a red or a green target to receive a reward. Reward was not guaranteed, but was instead governed by a random process based on the combination of preset weights (w) that were assigned to the ten shapes $\{w_1, w_2, \dots, w_{10}\}$. The sum of the four weights associated with the shapes shown in a trial established the log of the odds that reward would accompany a red or a green choice (see Methods Summary).

The large number of possible shape combinations (10^4 permutations; 715 unique combinations) prohibits memorization of specific four-shape patterns and encourages the monkeys to combine evidence from the probabilistic information conferred by each shape in each experimental trial. The optimal strategy is to choose the target with the larger reward probability. However, even when the odds favour red, for example, it is possible that green rather than red will be rewarded. We hoped that the sequential presentation of the shapes would permit a glimpse at how the brain combines probabilities to reach a decision.

Monkeys associated shapes with logLR

After extensive training (see Methods), the monkeys learned to base their decisions on the combined probabilities of the four shapes (Fig. 1b). They chose the red target when the evidence in favour of red was large, and chose the green target when the evidence in favour of green was large (large negative values in Fig. 1b). When the probability was between the two extremes, both monkeys chose either target, but tended to favour the one that was more likely to lead to reward.

To assess the degree to which each shape affected the monkeys' choices, we performed logistic regression (equation (7)), which models the log odds of a red choice as a sum of the weights assigned to each of the four shapes shown in each trial. Logistic regression thus provides a convenient estimate of the leverage that each of the ten shapes exerted on the monkeys' choices; this is given in units of log odds of producing a red choice. These weights, which we term the subjective weight of evidence (subjective WOE), bear a clear resemblance to the weights that were originally assigned to the ten shapes (Fig. 1c; $P < 10^{-5}$ for both monkeys, Spearman's rank correlation test). The rank ordering of the subjective WOE suggests that both monkeys learned the rules of the task, although neither monkey learned the weights perfectly (monkey J was clearly better than monkey H) (see Supplementary Fig. 6 for further analysis of the subjective WOE).

LIP responses are modulated by logLR

We recorded the activity of 64 neurons in the parietal lobe (area LIP) while the monkeys performed the probabilistic classification task. Many neurons in area LIP exhibit spatially selective persistent activity that reliably predicts whether an eye movement is planned into or out of the neuron's response field^{19–22}. The graded activity of these neurons is thought to represent a quantity used by the brain to make simple perceptual decisions^{8,9,11,12,23,24}. We hypothesized that these neural responses encode the logLR that the target in the neuron's response field is the one that will be rewarded.

To test this, we measured the firing rates from these neurons while the monkeys viewed the sequence of shapes that provided them with evidence in favour of reward at the red or green target. In each trial, either the red or the green target was in the neuron's response field (T_{in}). The other target (T_{out}) and the four shapes were placed outside the response field. The logLR is updated with the addition of each of each shape shown in the four presentation epochs. After the n th

¹Howard Hughes Medical Institute, Department of Physiology and Biophysics, National Primate Research Center, University of Washington, Box 357290, Seattle, Washington 98195–7290, USA.

shape (s_n), the WOE is:

$$\text{WOE} \equiv \log \text{LR}_n \equiv \log_{10} \frac{P(s_1, \dots, s_n | \text{reward at } T_{\text{in}})}{P(s_1, \dots, s_n | \text{reward at } T_{\text{out}})} \quad n = \{1, 2, 3 \text{ or } 4\} \quad (1)$$

where $P(x|y)$ denotes the probability of observing x , given that condition y holds true. If the red target is in the neuron's response field, then this value is roughly proportional to the sum of the weights assigned to the n shapes shown in the first n epochs; it equals exactly the sum of the weights once all four shapes are shown ($n = 4$) in the final epoch. The sign is reversed if the green target is in the response

field. In equation (1), one unit of WOE is called a 'ban'²⁵. We next show that LIP neurons modulate their firing rates as a function of these bans of evidence. Not surprisingly, LIP neurons were more responsive near the end of the trials when the monkeys chose T_{in} , as shown by the example neuron in Fig. 2a. The association between

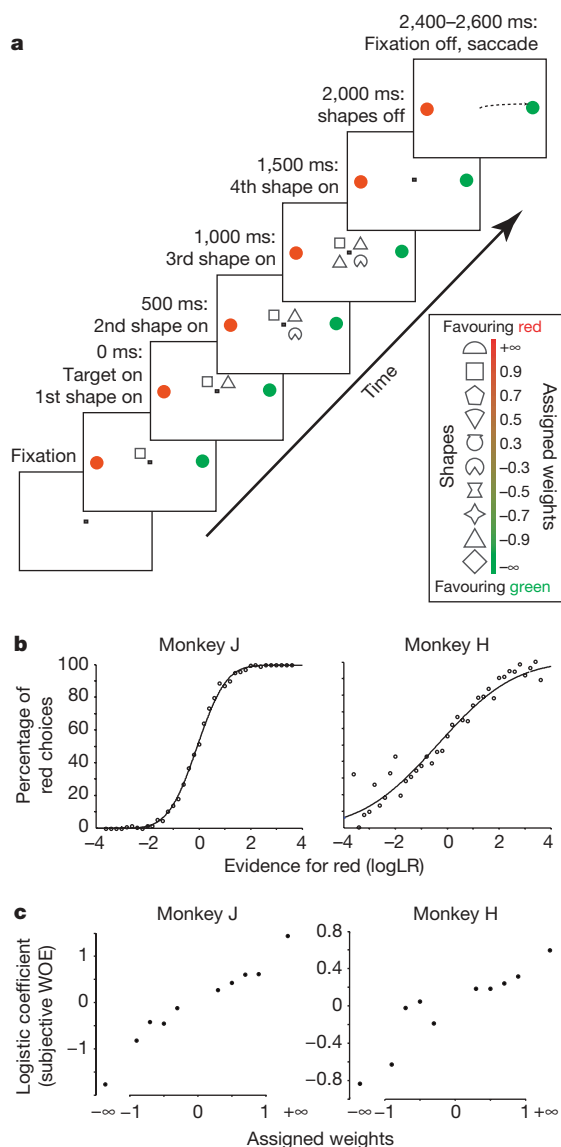


Figure 1 | Probabilistic categorization task. **a**, Task sequence. Four shapes were presented sequentially on the computer monitor near the centre of gaze. After a brief delay period, the monkey made an eye movement (saccade) to either the red or green choice target. During neural recording, one of the choice targets was in the response field of the neuron. The shapes were selected randomly in each trial from a larger set of ten (inset). The reward was determined probabilistically by summing the weights associated with the four shapes. The sum is the logarithm of the odds that the red target will be the one rewarded. **b**, Performance. The fraction of red choices is plotted as a function of the logLR conferred by the four shapes in favour of red. Curves are logistic fits to the data (equation (5)). Only trials that have finite logLR are included in this graph. **c**, Effect of individual shapes on choice. The leverage of each of the ten shapes on the probability of a red choice was inferred using logistic regression. This is the contribution that the shape has on the \log_{10} of the odds of a red choice. These values are plotted as a function of the assigned weights. Standard errors are smaller than the data points in **b** and **c**.

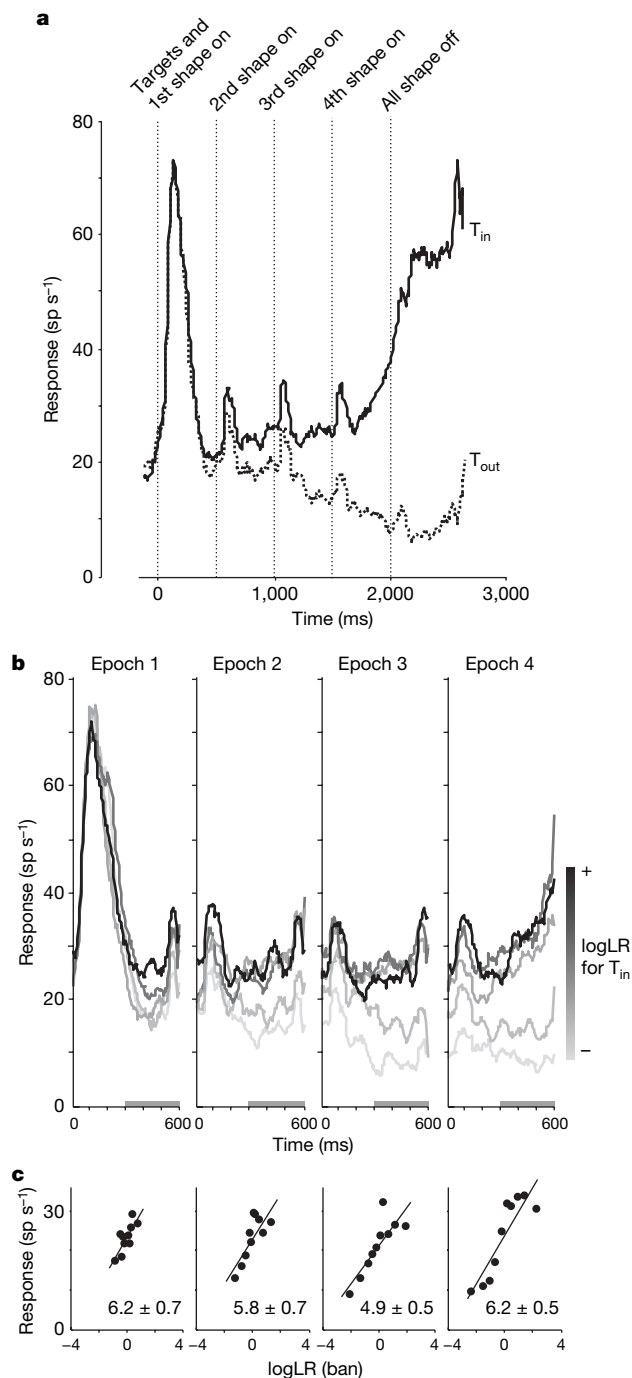


Figure 2 | Responses of an LIP neuron during probabilistic classification. **a**, Evolution of activity. The average firing rate is sorted by the monkeys' choices. The response averages reflect information from shapes shown in all four epochs. Response averages are drawn from 547 T_{in} and 575 T_{out} trials; bin width, 5 ms. **b**, Effect of logLR on firing rate. Response averages are aligned to the onset of the shapes and extend 100 ms into the subsequent epoch. The averages were computed for five quintiles of logLR in each epoch (indicated by shading). Note that a single trial contributes to only one curve in each epoch, but that the quintile can change depending on the sequence of shapes. **c**, Firing rate is affected by logLR. Average firing rate was calculated in the epochs indicated by the grey bars on the x axis of the graphs in **b**. The line is the weighted least-squares fit. The slope (\pm s.e.) provides a modulation index. Sample responses from this neuron can be viewed as Supplementary Movies 1–3.

firing rate and choice is a hallmark of LIP neurons and was nearly guaranteed by our sampling procedure (see Methods). In the present context, the LIP response reveals the outcome of a decision. Far more interesting is the evolution of the response that accompanied the sequential presentation of the shapes. Although a final decision must await the presence of all four shapes, the responses of this neuron were modulated by partial evidence as the shapes appeared in sequence.

The example neuron shown in Fig. 2a responded strongly to the onset of the choice targets that accompanied the first shape. This short latency response is probably caused by the appearance of T_{in} (ref. 26), but by the end of the first epoch the firing rate was affected by one of the ten shapes that appeared near the fixation point. The response was greater for shapes that provided evidence in favour of T_{in} (Fig. 2b). The response curves shown in this epoch sort the 20 possible conditions (10 shapes \times 2 choice target configurations) into quintiles that rank the total logLR for the choice target in the response field. (Note that the logLR quantifies the WOE in favour of T_{in} , regardless of its colour. We use this convention throughout the article for simplicity; but see Appendix C in Supplementary Information.) This difference was even more striking in the second epoch. Again, we grouped the 200 possible conditions (100 possible 2-shape sequences \times 2 choice target configurations) into 5 groups based on the total logLR for reward at T_{in} . The same analysis was performed in the next two epochs. In each epoch, the firing rates are affected by the logLR.

To quantify the effect of logLR on the neural response, we calculated the average firing rate for each trial in the epoch from 300 to

600 ms after shape onset, and plotted this value as a function of the logLR (Fig. 2c). The slope of the line of best fit provides an index of the response modulation by logLR and a test of statistical reliability (null hypothesis, H_0 : slope = 0). This neuron exhibited clear modulation of its firing rate as a function of logLR in all epochs ($P < 0.01$). The change in spike rate per ban is indicated in each panel. The positive value implies that the neuron increased its firing rate when the logLR favoured the target in its response field.

We observed a similar pattern of results for the sample of 64 neurons (Fig. 3). The response averages reveal a graded modulation of firing rates that correspond to the magnitude of logLR that favours the target in their response field. When the evidence was against the target in the response field, the population neuronal response decreased. The population average firing rate is well described by a linear function of logLR (Fig. 3b; $P < 0.01$; see equation (8) for H_0 : $b_n = 0$). The modulation indices (in units of spikes per second per ban) are shown for each neuron from the two monkeys (Fig. 3c). Although there is some heterogeneity across the population, the histograms reveal that the change in firing rate per ban is remarkably similar in all four epochs. This finding is supported by data from both monkeys (see Supplementary Fig. 5). In each epoch, LIP registers the appearance of a new shape by adjusting its firing rate to reflect the updated logLR in favour of T_{in} (see Supplementary Movies, which are described briefly in Supplementary Appendix D).

According to this theory, each of the ten shapes should cause a change in LIP activity in accordance with the weight it was assigned. We estimated these changes from the population by attempting to isolate the response to each new shape in each of the four epochs. The responses in the first epoch are obscured by the large visual response accompanying the onset of the choice targets. However, in epochs 2–4, we subtracted the firing rate that the neuron achieved in the previous epoch (see Methods). This procedure yields an estimate of the magnitude and time course of the change in firing rate caused by each of the ten stimuli (Fig. 4).

The shapes caused the firing rate to change with a fairly stereotypical time course, beginning ~ 150 –200 ms after shape onset without any obvious sign of decay (Fig. 4a). The change in firing rate appears to reflect both the sign and the magnitude of the assigned weights. This is easier to discern from Fig. 4b, which shows the average change in firing

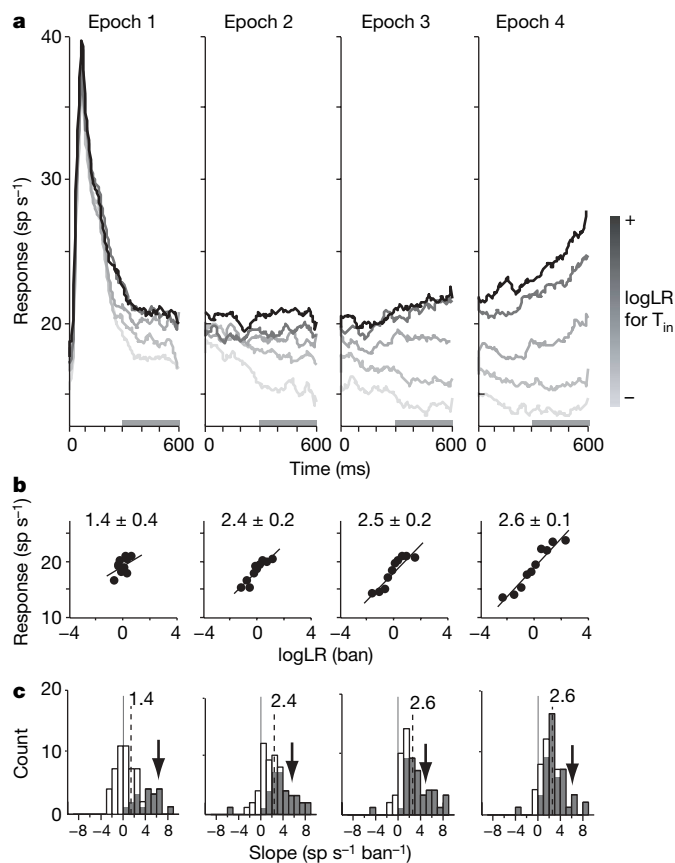


Figure 3 | Population analyses. **a**, Effect of logLR on the population average firing rate, presented in the same format as Fig. 2b (30,930 trials from 64 neurons; bin width, 5 ms). **b**, Firing rate varies linearly with logLR in each epoch, presented in the same format as Fig. 2c. The slope is the best fitting line to the population averages in **a**. **c**, Distribution of modulation indices from all 64 neurons. Indices were derived using the analysis in Fig. 2c. Slopes significantly different from zero are indicated by shaded bars ($P < 0.05$). Dashed lines represent the mean. Arrows indicate the indices for the example neuron in Fig. 2.

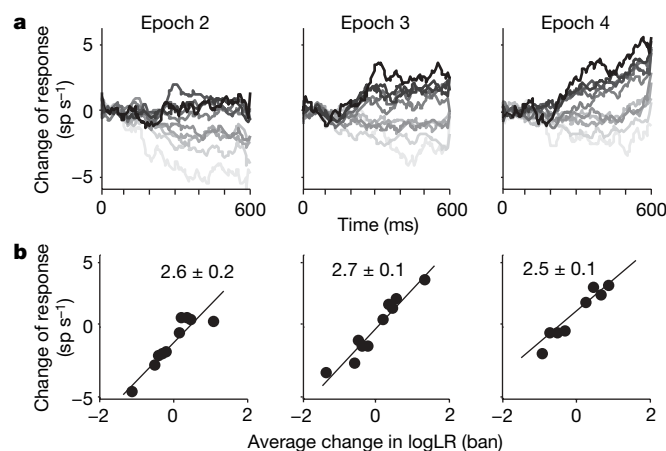


Figure 4 | Effect of individual shapes on LIP activity. The responses are computed by subtracting the firing rate preceding onset of the shape from the recorded value and then averaging the residual perturbations for each shape (see Methods). **a**, Time course of the change in firing rate caused by each of the ten shapes. The darkest curve corresponds to the shape that has the largest weight for T_{in} (semicircle shape when T_{in} is red; diamond shape when T_{in} is green); the lightest line corresponds to the shape that has the largest weight for T_{out} . **b**, The average change of firing rate induced by the ten shapes is plotted as a function of the change in logLR associated with the display of the shape in each of the epochs (Table 1). The two trump shapes are excluded in epoch 4.

rates in the epoch from 300 to 600 ms after shape onset, plotted as a function of logLR. Again, the change in firing rate is proportional to the change in logLR that accompanied addition of the shapes during the trial sequence. The change in spike rate per ban of evidence is similar to the values obtained from the total logLR (Fig. 3b). Furthermore, the changes in firing rate induced by the individual shapes are fairly similar in the three epochs depicted in Fig. 4.











Effect of logLR is not due to shape selectivity or eye movements

At face value, these results indicate that LIP neurons are capable of representing a quantity proportional to logLR. However, two important alternatives deserve consideration: first, LIP neurons might respond to the shapes themselves, independent of the probabilities they confer, and, second, the LIP might simply represent a commitment to move the eyes into or outside of the response field, regardless of the evidence that underlies that plan.

In principle, the observed changes in firing rates could be explained if LIP neurons responded better to some shapes than others^{27,28}. For example, such selectivity could arise as the monkeys learned that some stimuli are more predictive of reward than others^{11,24}. This idea can easily be dismissed. In this investigation, the colour of the target shown in the neuron's response field was randomized from trial to trial. Therefore, each shape combination could furnish evidence in favour either of or against the T_{in} target. In fact, the trials in the first and fifth quintiles in Figs 2b and 3a share the same shape combinations; the same is true for the second and fourth quintiles. Clearly, the responses are not determined by the shapes themselves, but by the bearing the shapes have on the pending decision.

A more serious concern is that the responses are associated with the monkeys' choice rather than with the evidence conferred by the shapes. After all, greater logLR in favour of T_{in} is associated with more neural activity and an increase in the likelihood of an eye movement to the response field. We wished to ascertain whether the effect of logLR on neural response is merely a reflection of the choice made at the end of the trial. In each epoch, we sorted the trials by the choice the monkey made (T_{in} or T_{out}) and subtracted the mean firing rate associated with that choice. The residual firing rates retain a strong dependency on the logLR (Fig. 5a; see also Supplementary Figs 7 and 8). Formally, this analysis is a simple extension of the linear regression depicted in Fig. 3b (see equation (9)). It allows us firmly to reject the hypothesis that the effect of logLR on firing rate is mediated by the eye movement choice ($P < 0.01$ in all epochs, equation (9), $H_0 : b_n = 0$). This is true even in the later epochs (T_{out} choices), in which the monkeys' eye movement choice clearly affects the responses. The same analysis performed on individual neurons also supports this conclusion: 84%, 76%, 60% and 28% of neurons in which logLR had significant effects on firing rate (epochs 1–4, respectively; filled bars in Fig. 3c) retained this significance when we controlled for choice ($P < 0.05$, equation (9)).

Table 1 | Change in logLR conferred by the ten shapes in each epoch

Shapes	$\Delta \log LR$ (ban)			
	Epoch I	Epoch II	Epoch III	Epoch IV
 	± 1.09	± 1.31	± 1.67	–
 	± 0.32	± 0.37	± 0.49	± 0.9
 	± 0.25	± 0.30	± 0.39	± 0.7
 	± 0.18	± 0.21	± 0.28	± 0.5
 	± 0.11	± 0.12	± 0.17	± 0.3

Values for epochs II–IV are averages, which take into account all possible shape combinations in the preceding epochs. The sign of the change depends on whether the red or green target is in the neuron's response field. The change in logLR is not meaningful in epoch 4 for the trump shapes because the majority of combinations result in $\pm \infty$. For the same shapes, there are 18 combinations in epoch III that result in $\pm \infty$; these are excluded from this average.

The preceding analysis indicates that the monkeys' choice does not explain the effect of logLR on firing rate, but it is possible that the monkey favours one choice or the other in each epoch and simply changes this binary commitment in the subsequent epoch upon acquisition of new information. This idea could explain the weak effect of choice on neural response in early epochs while leaving open the possibility that logLR is not in fact represented. According to this formulation, the intermediate levels of activity that appear to correspond to different levels of logLR are actually mixtures of responses associated with just two 'states' of commitment, albeit temporary.

To test for this possibility, we measured the variance in firing rates across trials. If intermediate levels of firing are mixtures of high and low firing rate states, then the variance should be predicted by sampling responses from the distributions of responses associated with the highest and lowest quintiles (grey curves, Fig. 5b; see Methods). In contrast, if the intermediate firing rates represent actual firing rate modes, the variance should be approximately proportional to the mean firing rate^{29–31}, corresponding to straight lines in Fig. 5b. This analysis firmly rejects the conjecture that the intermediate firing rates are actually mixtures of trials drawn from the extreme quintiles (H_0 : mixture; $P < 0.01$ in the four shape epochs, F -test). It provides clear evidence that the intermediate firing rates represent a graded quantity that is proportional to logLR. In fact, a significant deviation from a graded representation of logLR (straight lines in Fig. 5b) is only observed in the fourth epoch and in the delay period (H_0 : non-mixture; $P < 0.01$), when a categorical decision is expected.

Together, the analyses in Fig. 5 indicate that the LIP activity represents both the outcome of the decision process—an eye movement—and the accumulated evidence on which this decision is based. The evidence in the form of logLR affects both LIP firing rates and the choice. In early epochs the firing rates are predominantly affected by the logLR, whereas in later epochs the rates are associated with both logLR and choice. Note, however, that the effect of logLR on firing rate is nearly the same in all epochs (Fig. 3b); it is also the same for the individual shapes (Fig. 4b). Therefore, even in the later epochs when

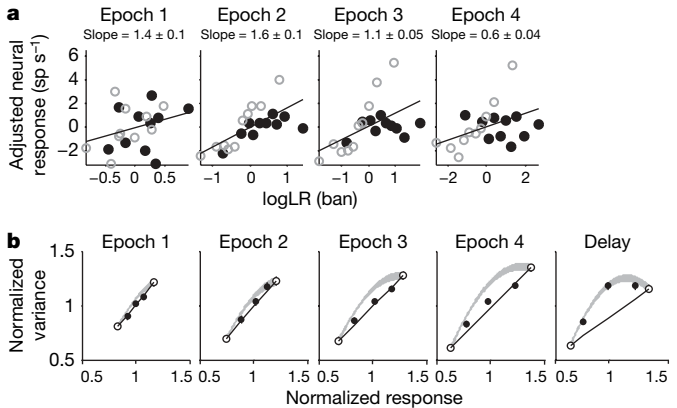


Figure 5 | The effect of logLR on firing rate is not explained by motor planning. **a**, Analysis of the potential confounding effect of eye movement. Data points are mean values of adjusted firing rates, detrended for the effect of eye movement at the end of the trial. Detrending was achieved for each cell and in each epoch by subtracting the mean firing rate for T_{in} or T_{out} choices from the firing rate in single trials. Lines are based on fits to equation (9) (slope = b_n) using all trials in each epoch ($N \approx 21,000$ to 30,000). Slopes (\pm s.e.) are in units of spikes per second per ban. For the purpose of display, points represent ten divisions of logLR ($\sim 1,500$ trials per point). Filled circles, T_{in} choices; open circles, T_{out} choices. **b**, Rejection of a mixture model. Graphs show the relationship between the mean firing rate and variance for five quintiles of logLR in a 300 ms window in each epoch (see Methods) and in the delay period (300 ms presaccadic interval). The values are normalized for each neuron and are averaged to produce the points shown (error bars are s.e.m.; $n = 64$ neurons). The grey line shows the predicted variance (\pm s.e.) if the intermediate levels of firing rate were composed of mixtures of firing rate states associated with the lowest and highest quintiles (open symbols).

the firing rates are dominated by the monkeys' ultimate choice (Supplementary Fig. 7), parsimony suggests that logLR causes a change in LIP firing rate, which in turn causes (or is associated with) a change in the probability of a T_{in} choice.

LIP reflects behavioural variations not explained by logLR

To date, we have focused our analyses on the relationship between neural response and the actual logLR conferred by the ten shapes. Although the monkeys learned these probabilistic relationships, they were not perfect. As shown in Fig. 2, the leverage that the ten shapes exerted on behaviour, termed the subjective WOE, was different between the two monkeys and exhibited some departures from the scaling imposed by the assigned weights. We wondered whether LIP activity is correlated with these idiosyncrasies. To test this, we conducted a nested regression analysis, which was designed to reveal whether knowledge of the discrepancies between logLR and the subjective WOE could explain some of the remaining variance in neural responses (see Methods). For all four epochs tested in both monkeys, knowledge of subjective WOE improved the regression significantly ($P < 0.005$, equation (10), $H_0: d_n = 0$). We conclude that LIP neurons encode the systematic biases formed by the monkeys in the association of shape and evidence bearing on the decision.

In addition to these systematic biases, the LIP recordings help to explain why the monkeys vary their choices even when faced with the same evidence (for example, Supplementary Fig. 2a). The variation in firing rate from a single neuron has a detectable effect on the choices the monkey makes: on average, a variation of 1 spike per second from a single neuron was equivalent to ~ 0.1 ban of evidence ($P < 0.05$ for all but the first epoch for both monkeys; see equations (11) and (12)). Such correlations are thought to arise because the variability in firing rate from single neurons is shared to a small degree with other neurons that underlie the choice^{31–33}. Thus, the mechanism underlying probabilistic behaviour in our task resembles the mechanism underlying errors on difficult perceptual tasks when evidence is weak—that is, signals that affect behaviour are often contaminated by noise.

Discussion

Humans make probabilistic classifications in a variety of settings in which information bears on—but does not guarantee—an outcome. In the weather-prediction task, for example, a human is asked to predict the weather based on a set of cards with symbols on them. Like our shapes, these symbols possess probabilistic associations with the possible outcomes (rain or sunshine), and the subject receives feedback on his or her decision. This task is thought to rely on procedural memory circuits that involve the basal ganglia during learning³⁴. Although we cannot be sure that the monkeys in this investigation use the same strategy as humans, this study begins to raise questions about the connection between so-called habit learning and cortical computations associated with probabilistic inference. The capacity to train monkeys on this task opens new possibilities for further investigations.

As has been found in other decision-making tasks, when a decision is ultimately expressed through a movement, neurons in high-level command structures of the brain mediate the logical connection between evidence and plan of action^{8,11,23,35–38}. As in simple decisions, this evolution of the LIP response represents the accumulation of evidence leading to commitment to a behavioural choice³⁹. Indeed, it has been suggested that the build-up of activity in the brain associated with a wide variety of simple decisions is an approximation to the summation of logarithms of probabilities, likelihoods, odds, and so on^{16,40–44}. The present findings provide direct confirmation of these theoretical insights for the first time. They demonstrate directly that the firing rates of LIP neurons are proportional to the logLR conferred by the shape stimuli in our task.

This interpretation does not exclude the possibility that LIP encodes many other useful quantities for decision-making (for example, value, utility or reward expectation^{11,24,45}), especially if the encoding is in

units roughly proportional to logLR⁴⁶. Even in this study, there are alternatives to the logLR that would produce similar results. For example, the brain might approximate the logLR associated with each shape on the assumption of conditional independence—in essence, that the logLR for epoch 1 holds for all epochs, what we term 'naive WOE'. This quantity would be easy to estimate from the frequency of observing each shape in association with reward at the red and green targets. It has the virtue of simplicity, and it predicts the pattern of response seen with the two 'trump' shapes (half-circle and diamond; see Supplementary Appendix A). A related alternative is the subjective WOE—the quantity we derived from our behavioural analysis (Fig. 1c). Not only is subjective WOE approximately proportional to the logLR but also it reflects the systematic error in the monkeys' choices. Indeed, we would have obtained similar results in all analyses in this article if we substituted subjective WOE for logLR. We presented the results of behavioural measurements and neural recordings as a function of logLR because that is the experimentally controlled independent variable. Thus, we have established a link between logLR and both neural activity and choice.

We do not know how the brain converts information about shape into a number proportional to logLR, and we do not know whether area LIP plays a role in this transformation or whether it represents only the outcome of this conversion. Presumably, neurons in ventral stream visual areas play a role in discriminating the shapes⁴⁷, but how their responses are converted to a logLR value is not known. In a motion-discrimination task it has been shown that LIP neurons represent the accumulated difference in firing rates from direction-selective neurons with opposing direction preferences^{13,48}. This difference approximates a logLR^{43,49}. It remains to be seen whether a similar readout of ventral stream neurons occurs in our probabilistic classification task.

Although designed to test a straightforward hypothesis about a particular representation of probability, the present study exposes the brain's capacity to extract probabilistic information from a set of symbols and to combine this information over time. Thus, we have demonstrated a crude capacity for probabilistic inference in monkeys—a capacity that might underlie cognitive reasoning in humans.

METHODS SUMMARY

Two monkeys (*Macaca mulatta*) were trained in a probabilistic categorization task (Fig. 1). In each trial, the monkey maintained its gaze at a fixation point. Two choice targets (red and green) were displayed in opposite hemifields (eccentricity range 8° to 12°). Next, four shapes were presented at regular 500 ms intervals. The four shape positions (0.5° from the fixation point) were displayed in a random order in each trial. The shapes were erased 500 ms after onset of the fourth. A memory delay period (450–550 ms) preceded offset of the fixation point, which instructed the monkey to make an eye movement to one of the choice targets. A liquid reward was administered probabilistically for either the red or green choice, as described below.

Each shape was drawn randomly from a pool of ten shapes (sampled with replacement). We assigned each shape a unique weight: $\{w_1, w_2, \dots, w_{10}\} = \{-\infty, -0.9, -0.7, -0.5, -0.3, +0.3, +0.5, +0.7, +0.9, +\infty\}$. The extreme weights, which we term 'trumps', can be regarded as a pair of very large numbers, $w_1 = \lim_{x \rightarrow \infty} (-x)$, $w_{10} = \lim_{x \rightarrow \infty} (x)$, which cancel when they appear together. The sum of the weights associated with the four shapes governed the probability that a red (R) or a green (G) choice would be rewarded.

$$P(R|s_1, s_2, s_3, s_4) = \frac{\sum_{i=1}^4 w_i}{1 + \sum_{i=1}^4 w_i} \quad (2)$$

$$P(G|s_1, s_2, s_3, s_4) = 1 - P(R|s_1, s_2, s_3, s_4)$$

where s_i represents the shape shown in the i th epoch. The sum of weights is the log of the posterior odds in favour of red.

$$\log_{10} \frac{P(R|s_1, s_2, s_3, s_4)}{P(G|s_1, s_2, s_3, s_4)} = \sum_{i=1}^4 w_i \quad (3)$$

The computer controlled the reward using $P(R|s_1, s_2, s_3, s_4)$.

We recorded extracellularly from single neurons in LIP while the monkeys performed the task with one of the choice targets (T_{in}) in the neuron's response field. Neither T_{out} nor the shapes were near the neuron's response field. We selected all well isolated neurons that had spatially selective persistent activity during a memory delay period⁵⁰. A total of 64 neurons were recorded for at least 100 trials and were included in our analysis (45 and 19 from monkeys J and H, respectively). See Methods for further details.

All experimental procedures were performed in accordance with the NIH Guide for the Care and Use of Laboratory Animals and were approved by the University of Washington Animal Care Committee.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 30 November 2006; accepted 18 April 2007.

Published online 3 June 2007.

- Schall, J. D. Neural correlates of decision processes: neural and mental chronometry. *Curr. Opin. Neurobiol.* **13**, 182–186 (2003).
- Glimcher, P. W. The neurobiology of visual-saccadic decision making. *Annu. Rev. Neurosci.* **26**, 133–179 (2003).
- Horowitz, G. D., Batista, A. P. & Newsome, W. T. Representation of an abstract perceptual decision in macaque superior colliculus. *J. Neurophysiol.* **91**, 2281–2296 (2004).
- Gold, J. I. & Shadlen, M. N. Banburismus and the brain: decoding the relationship between sensory stimuli, decisions, and reward. *Neuron* **36**, 299–308 (2002).
- Romo, R., Hernandez, A. & Zainos, A. Neuronal correlates of a perceptual decision in ventral premotor cortex. *Neuron* **41**, 165–173 (2004).
- Barlow, H. Conditions for versatile learning. Helmholtz's unconscious inference, and the task of perception. *Vision Res.* **30**, 1561–1571 (1990).
- Gold, J. I. & Shadlen, M. N. The influence of behavioral context on the representation of a perceptual decision in developing oculomotor commands. *J. Neurosci.* **23**, 632–651 (2003).
- Roitman, J. D. & Shadlen, M. N. Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *J. Neurosci.* **22**, 9475–9489 (2002).
- Shadlen, M. N. & Newsome, W. T. Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *J. Neurophysiol.* **86**, 1916–1936 (2001).
- Kim, J. N. & Shadlen, M. N. Neural correlates of a decision in the dorsolateral prefrontal cortex of the macaque. *Nature Neurosci.* **2**, 176–185 (1999).
- Platt, M. L. & Glimcher, P. W. Neural correlates of decision variables in parietal cortex. *Nature* **400**, 233–238 (1999).
- Hanks, T. D., Ditterich, J. & Shadlen, M. N. Microstimulation of macaque area LIP affects decision-making in a motion discrimination task. *Nature Neurosci.* **9**, 682–689 (2006).
- Huk, A. C. & Shadlen, M. N. Neural activity in macaque parietal cortex reflects temporal integration of visual motion signals during perceptual decision making. *J. Neurosci.* **25**, 10420–10436 (2005).
- Link, S. W. & Heath, R. A. A sequential theory of psychological discrimination. *Psychometrika* **40**, 77–105 (1975).
- Wald, A. *Sequential Analysis* (Wiley, New York, 1947).
- Gold, J. I. & Shadlen, M. N. The neural basis of decision making. *Annu. Rev. Neurosci.* **30**, 535–574 (2007).
- Knowlton, B. J., Mangels, J. A. & Squire, L. R. A neostriatal habit learning system in humans. *Science* **273**, 1399–1402 (1996).
- Gluck, M. A., Shohamy, D. & Myers, C. How do people solve the “weather prediction” task?: individual variability in strategies for probabilistic category learning. *Learn. Mem.* **9**, 408–418 (2002).
- Gnadt, J. W. & Andersen, R. A. Memory related motor planning activity in posterior parietal cortex of monkey. *Exp. Brain Res.* **70**, 216–220 (1988).
- Platt, M. L. & Glimcher, P. W. Responses of intraparietal neurons to saccadic targets and visual distractors. *J. Neurophysiol.* **78**, 1574–1589 (1997).
- Colby, C. L. & Goldberg, M. E. Space and attention in parietal cortex. *Annu. Rev. Neurosci.* **22**, 319–349 (1999).
- Ipata, A. E., Gee, A. L., Goldberg, M. E. & Bisley, J. W. Activity in the lateral intraparietal area predicts the goal and latency of saccades in a free-viewing visual search task. *J. Neurosci.* **26**, 3656–3661 (2006).
- Leon, M. I. & Shadlen, M. N. Representation of time by neurons in the posterior parietal cortex of the macaque. *Neuron* **38**, 317–327 (2003).
- Sugrue, L. P., Corrado, G. S. & Newsome, W. T. Matching behavior and the representation of value in the parietal cortex. *Science* **304**, 1782–1787 (2004).
- Good, I. J. Studies in the history of probability and statistics. XXXVII A.M. Turing's statistical work in World War II. *Biometrika* **66**, 393–396 (1979).
- Bisley, J. W., Krishna, B. S. & Goldberg, M. E. A rapid and precise on-response in posterior parietal cortex. *J. Neurosci.* **24**, 1833–1838 (2004).
- Sereno, A. B. & Maunsell, J. H. Shape selectivity in primate lateral intraparietal cortex. *Nature* **395**, 500–503 (1998).
- Sereno, A. B. & Amador, S. C. Attention and memory-related responses of neurons in the lateral intraparietal area during spatial and shape-delayed match-to-sample tasks. *J. Neurophysiol.* **95**, 1078–1098 (2006).
- Tolhurst, D. J., Movshon, J. A. & Dean, A. F. The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision Res.* **23**, 775–785 (1983).
- Geisler, W. S. & Albrecht, D. G. Visual cortex neurons in monkeys and cats: detection, discrimination and identification. 897–920 (1997).
- Shadlen, M. N. & Newsome, W. T. The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *J. Neurosci.* **18**, 3870–3896 (1998).
- Zohary, E., Shadlen, M. N. & Newsome, W. T. Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature* **370**, 140–143 (1994); erratum **371**, 6495 (1994).
- Parker, A. J. & Newsome, W. T. Sense and the single neuron: probing the physiology of perception. *Annu. Rev. Neurosci.* **21**, 227–277 (1998).
- Knowlton, B. J., Squire, L. R. & Gluck, M. A. Probabilistic classification learning in amnesia. *Learn. Mem.* **1**, 106–120 (1994).
- Gold, J. I. & Shadlen, M. N. Representation of a perceptual decision in developing oculomotor commands. *Nature* **404**, 390–394 (2000).
- Cisek, P. & Kalaska, J. F. Neural correlates of reaching decisions in dorsal premotor cortex: specification of multiple direction choices and final selection of action. *Neuron* **45**, 801–814 (2005).
- Wise, S. P. & Murray, E. A. Role of the hippocampal system in conditional motor learning: mapping antecedents to action. *Hippocampus* **9**, 101–117 (1999).
- Sugrue, L. P., Corrado, G. S. & Newsome, W. T. Choosing the greater of two goods: neural currencies for valuation and decision making. *Nature Rev. Neurosci.* **6**, 363–375 (2005).
- Smith, P. L. & Ratcliff, R. Psychology and neurobiology of simple decisions. *Trends Neurosci.* **27**, 161–168 (2004).
- Jazayeri, M. & Movshon, J. A. Optimal representation of sensory information by neural populations. *Nature Neurosci.* **9**, 690–696 (2006).
- Carpenter, R. & Williams, M. Neural computation of log likelihood in control of saccadic eye movements. *Nature* **377**, 59–62 (1995).
- Holmes, P. et al. Optimal decisions: from neural spikes, through stochastic differential equations, to behavior. *IEICE Trans. Fundamentals* **88**, 2496–2503 (2005).
- Gold, J. I. & Shadlen, M. N. Neural computations that underlie decisions about sensory stimuli. *Trends Cogn. Sci.* **5**, 10–16 (2001).
- Ma, W. J., Beck, J. M., Latham, P. E. & Pouget, A. Bayesian inference with probabilistic population codes. *Nature Neurosci.* **9**, 1432–1438 (2006).
- Dorris, M. C. & Glimcher, P. W. Activity in posterior parietal cortex is correlated with the relative subjective desirability of action. *Neuron* **44**, 365–378 (2004).
- Green, D. M. & Swets, J. A. *Signal Detection Theory and Psychophysics* (John Wiley and Sons, New York, 1966).
- Tanaka, K. Inferotemporal cortex and object vision. *Annu. Rev. Neurosci.* **19**, 109–139 (1996).
- Ditterich, J., Mazurek, M. & Shadlen, M. N. Microstimulation of visual cortex affects the speed of perceptual decisions. *Nature Neurosci.* **6**, 891–898 (2003).
- Shadlen, M. N., Hanks, T. D., Churchland, A. K., Kiani, R. & Yang, T. in *Bayesian Brain: Probabilistic Approaches to Neural Coding* (eds Doya, K., Ishii, S., Rao, R. & Pouget, A.) 209–237 (MIT Press, Cambridge, 2006).
- Hikosaka, O. & Wurtz, R. H. Visual and oculomotor functions of monkey substantia nigra pars reticulata. III. Memory-contingent visual and saccade responses. *J. Neurophysiol.* **49**, 1268–1284 (1983).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature. A summary figure is also included.

Acknowledgements We thank H. Brew, A. Churchland, T. Hanks, R. Kiani and J. Palmer for advice and comments, M. Mihali and V. Skypceck for technical assistance, and M. McKinley for preparing movie demonstrations. This work was supported by the Howard Hughes Medical Institute (HHMI) and grants from the NEI and NCRR.

Author Contributions The authors designed the project together. T.Y. collected data and performed the data analysis. T.Y. and M.N.S. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to T.Y. (tianming@u.washington.edu) or M.N.S. (shadlen@u.washington.edu).

METHODS

Here we describe important details about the behavioral and physiological methods used in this study. We also explain the analyses described in the main text. **Task.** In each trial, the monkey maintained its gaze within $\pm 1^\circ$ of a fixation point (0.2° in diameter) centred on the video monitor. After a variable delay, two choice targets (red and green) were displayed in opposite hemifields (eccentricity range 8° to 12° ; median 10°). At the same time (monkey J) or 500 ms later (monkey H), the first of four shape-stimuli appeared near the fixation point. Next, shapes 2, 3 and 4 were presented at regular 500 ms intervals. Each shape was randomly drawn from a pool of ten shapes (probability 0.1; sampled with replacement). The four positions (0.5° from the fixation point) were assigned in random order in each trial. The shapes were drawn as white lines on a black background (line length 2° ; thumb print $<(0.5^\circ)^2$; Fig. 1). The shapes were erased 500 ms after onset of the fourth. A memory delay period (450–550 ms) preceded the removal of the fixation point, which instructed the monkey to make an eye movement to one of the choice targets. A liquid reward was administered probabilistically for either the red or the green choice, as described below.

Training. Animals were seated comfortably in a primate chair with their head fixed. Eye position was monitored using a scleral search coil (CNC Engineering)⁵¹. Horizontal and vertical eye positions were sampled at 500 Hz using the NIH Rex system⁵². Both animals had received prior training on delayed eye-movement tasks. We began training with a one-shape version of the categorization task. We first used the shapes with $w = \pm \infty$, for which the 'correct' choice would yield reward with 100% certainty. Once animals consistently achieved at least 70% correct choices, we enlarged the pool of possible test shapes using the pair of shapes that had the next largest weights. Reward was administered probabilistically for these new shapes (using a variant of equation (2) with just one term in the sum), but we kept track of the fraction of trials in which the monkey chose the target that was favoured by the weights, even if the choice did not pay off. When the monkey reached 70% correct choices, we added the third pair of shapes, and so on. Here, correct means the choice of the target that has larger reward probability.

Once the monkey learned this one-shape version of the task (ten and four weeks for monkeys J and H, respectively), we began training on the n -shape version of the task. We provided only minimal exposure to versions of the task that had two and three shapes (two and one weeks for monkeys J and H, respectively), because, to our surprise, the monkeys were willing to perform the four-shape version without these intervening steps. Monkey J underwent 8 weeks of training in the full four-shape version of the task ($\sim 130,000$ trials). By the first recording session, this monkey was choosing the better reward target in more than 85% of trials. Monkey H received 10 weeks of training on the full 4-shape version of the task ($\sim 134,000$ trials). Because a veterinary complication (endometriosis) favoured an expedited testing schedule, we started recording when this monkey chose the better reward target in $\sim 75\%$ of trials.

Recording. We recorded extracellularly from single neurons in area LIP while the monkeys performed the task. We targeted neurons in the posterior third of the ventral division of LIP^{53,54} by registering recording locations estimated from our magnetic resonance imaging studies (Supplementary Fig. 4a) to a high resolution scan supplied with the Caret atlas⁵⁵ (<http://brainmap.wustl.edu/caret>). A hydraulic microdrive mounted over the cylinder advanced the electrodes to the desired depth (FHC, <http://www.fh-co.com/>). Spikes from individual neurons were isolated using a dual voltage–time window discriminator (Bak Electronics). They were recorded with a resolution of 1 ms for off-line analysis using custom software written in Matlab (Mathworks). We screened neurons using a simple delayed eye-movement paradigm^{50,56}. We selected for further study all well isolated neurons that had spatially selective persistent activity during the memory delay period (500–1,000 ms) between target flash and an eye movement to its remembered location ($\pm 2.5^\circ$) (ref. 50). Neurons that have this property are abundant in the ventral division of LIP⁵⁶.

Data analysis. The main independent variable in this study is the WOE conferred by the sequence of shapes ($\log LR_m$, equation (1)). Bayes' rule ensures that this is equal to the log of the posterior odds (equation (3)) once all four shapes are shown. This is because the prior probability of reward at red (R) is equal to the prior probability of reward at green (G):

$$\log_{10} \frac{P(s_1, s_2, s_3, s_4 | R)}{P(s_1, s_2, s_3, s_4 | G)} = \log_{10} \frac{P(R | s_1, s_2, s_3, s_4)}{P(G | s_1, s_2, s_3, s_4)} = \sum_{i=1}^4 w_i \quad (4)$$

The $\log LR$ from partial evidence (that is, fewer than four shapes) was calculated by tabulating the expected frequencies of reward associated with each shape combination, rather than the partial sums of weights. For an explanation of the difference between these two approaches, see Supplementary Appendices A and B. To calculate the $\log LR$ associated with a particular shape in each epoch

(Table 1), we averaged the change in $\log LR$ conferred by the shape; this was done for all possible shape combinations in the preceding epochs, excluding the appearance of the trump shapes when they guaranteed reward at red or green (that is, an unbalanced trump in the fourth epoch and the rare run of two or more unbalanced trumps in epoch three).

Analyses of behavioural data. We used a variety of logistic models to ascertain the effect of the shape combinations on choice. The sigmoid in Fig. 1b is described by

$$P_{\text{red}} = \frac{10^Q}{1 + 10^Q} \quad \text{where } Q = \beta_0 + \beta_1 \log_{10} \frac{P(s_1, s_2, s_3, s_4 | R)}{P(s_1, s_2, s_3, s_4 | G)} \quad (5)$$

where the β_i are fitted coefficients (method of maximum likelihood assuming the Bernoulli distribution of binary choices). Nearly identical results are obtained by replacing the $\log LR$ in equation (5) by the sum of the four weights or by an alternative to the $\log LR$, termed naive WOE, described in the Appendix A of the Supplementary Information. We use equation (5) for consistency with other analyses for which WOE (as defined in equation (1)) is the only mathematically correct option (see Supplementary Appendix A). The trials that used infinite total weights were not included in this analysis. In these trials, we assessed the leverage of shapes that have finite weight on choice by letting

$$P_{\text{red}} = \frac{10^Q}{1 + 10^Q} \quad \text{where } Q = \gamma_0 + \gamma_1 \sum_{i=1}^3 w_i \quad (6)$$

where the w_i are the weights assigned to the three shapes that have finite weight. We performed this test separately for the two trump shapes. The null hypothesis is that the three shapes that have finite weight do not affect choice in the presence of a trump shape (Supplementary Fig. 2b).

To estimate the effect of individual shapes on the monkeys' choices, we incorporated a term for each of the ten possible shapes in the logistic regression:

$$P_{\text{red}} = \frac{10^{Q^*}}{1 + 10^{Q^*}} \quad \text{where } Q^* = \sum_{j=1}^{10} w_j^* N_j \quad (7)$$

and the N_j are the counts for each shape type presented in a trial. The ten fitted coefficients w_j^* are the subjective WOE. Subscript j in this equation refers to shape (see also Supplementary Fig. 6).

Analyses of physiological data. Our primary analyses test the relationship between the $\log LR$ associated with shape sequences and the neuronal firing rate. For most analyses, firing rates were estimated in single trials in each of the four shape epochs from 300–600 ms after shape onset. Responses tended to be stable in this interval and were uncontaminated by onset of choice targets (first epoch) or the subsequent shape (epochs two and three). None of the results in this article rest critically on the definition of this epoch. Although many graphs display data grouped by $\log LR$ (quintiles, running means and binned data), all fits and statistical tests were performed using individual trials rather than mean responses, unless otherwise noted. Fits to individual trials used least squares, and fits to means (and their s.e.) used weighted least squares.

For the analyses in Figs 2 and 3, we used simple linear regression in each epoch:

$$y_n = a_n + b_n \log LR_n \quad (8)$$

where y_n is the firing rate (in each trial) in epoch n , and $\log LR_n$ is given by equation (1). The fitted coefficients, a and b , are the intercept and slope of the line, respectively. We refer to the latter as a modulation index for the neuron or the population.

To test whether the effect of $\log LR$ is explained by the eye movement made at the end of the trial, we added a third term to this equation:

$$y_n = a_n + b_n \log LR_n + c_n I_{\text{eye}}, \quad I_{\text{eye}} = \begin{cases} 1 & \text{if eye movement is to } T_{\text{in}} \\ 0 & \text{if eye movement is to } T_{\text{out}} \end{cases} \quad (9)$$

The null hypothesis is that the effect of $\log LR$ on LIP response is explained by the eye movement ($H_0: b_n = 0$).

We also considered the possibility that firing rates might be explained better by the weights that the monkeys gave to the shapes. To test this, we added a term to equation (9) to represent the difference between the assigned $\log LR$ and the subjective WOE:

$$y_n = a_n + b_n \log LR_n + c_n I_{\text{eye}} + d_n \Delta_n \quad \text{where } \Delta_n = \sum_{i=1}^n \text{subjective WOE}_i - \log LR_n \quad (10)$$

The null hypothesis, $d_n = 0$, asserts that the subjective WOE affects the LIP firing rate in a manner that is explained by the $\log LR$ assigned to the shapes.

To estimate the effect of individual shapes on LIP responses (Fig. 4), we isolated the change in firing rate from the 'baseline' level attained prior to presentation of the shape. The baseline firing rate was estimated from individual

trials using the average response over the 200 ms preceding the onset of the shape (that is, from the preceding epoch). We subtracted this baseline from the response after the onset of the shape, and averaged the resulting traces across the population of 64 neurons.

To determine whether intermediate levels of firing rate could be explained as mixtures of high and low firing rate states (Fig. 5b), in each epoch, we grouped the neural responses into quintiles based on the rank of logLR. For each cell in each epoch, we calculated five mean firing rates and their associated variances. In the mixture model, we assumed that the responses comprising the three intermediate quintiles represent mixtures of values drawn from the extreme quintiles, termed anchors, representing high and low states. The proportion of high and low samples in each of the mixtures was set to match the mean firing rate for each bin. We then calculated the expected variance based on these proportions (theoretical curves in Fig. 5b). In the non-mixture model, the variance was established by interpolating between the variances associated with the anchor distributions. This is equivalent to assuming a constant Fano factor for all the quintiles, consistent with the idea that the intermediate mean is not a mixture. For the graphs in Fig. 5b, we normalized the variances (from model and data) to the mean of the five sample variances, and we normalized the mean firing rates to the average of the five means. This places the point for the middle quintile near [1,1] on these plots. We then averaged these normalized values over all neurons. We used F statistics to compare the observed variance to the value predicted from the mixture model and to the predicted value from the non-mixture model. The P values reported in association with Fig. 5b are based on combined data from all neurons; data from single neurons support the trends in the figure but lack the power to reject mixture or non-mixture, or both. Note that this analysis is more sensitive than standard tests for bimodality. For example, Hartigan's dip test failed to reject unimodality of the distributions associated with intermediate quintiles in all epochs, including the delay period ($P > 0.9$, combined standardized responses).

Correlation between choice and neural response. To measure the relationship between trial-by-trial variation in LIP response and behavioural choice, we incorporated firing rate as an additional term in the logistic model that explains the monkeys' choices based on the shape combinations. Using a nested models approach, we begin with the successful account of behaviour using the definition of logLR in units referable to the target in the neuron's response field:

$$P_{\text{RF}} = \frac{10^Q}{1 + 10^Q} \quad \text{where } Q = \beta_0 + \beta_1 \log_{10} \frac{P(s_1, s_2, s_3, s_4 | \text{reward at } T_{\text{in}})}{P(s_1, s_2, s_3, s_4 | \text{reward at } T_{\text{out}})} \quad (11)$$

This would produce the same graph as Fig. 1b with a new ordinate, labelled 'Probability of T_{in} choice'. In each epoch, $n = 1$ to 4, we incorporated a term that reflects the firing rate of the neuron. This value is the residual error after fitting firing rate as a function of logLR and subjective WOE in the n th epoch, which we incorporated into an extended logistic model of choice:

$$P'_{\text{RF}} = \frac{10^{Q'}}{1 + 10^{Q'}} \quad \text{where } Q' = Q + \beta_2 r_n \quad (12)$$

where r_n is the residual error in the fit to equation (10) but without the eye movement term (that is, $c_n = 0$). The ratio β_2/β_1 in equations (11) and (12) furnishes an estimate of the effect that the variable discharge of a single neuron in a single trial has on choice, in units of bans per spikes per second, that cannot be explained by the stimulus.

Unless otherwise noted, fits were performed using the method of maximum likelihood. Standard errors on the estimated parameters (coefficients) were obtained using the Hessian matrix of the log likelihood. The parameter estimates and their s.e. values were used to construct t statistics to test the null hypotheses.

51. Robinson, D. A. A method of measuring eye movement using a scleral search coil in a magnetic field. *IEEE Trans. Biomed. Eng.* **10**, 137–145 (1963).
52. Hayes, A. V., Richmond, B. J. & Optician, L. M. A UNIX-based multiple process system for real-time data acquisition and control. *WESCON Conf. Proc.* **2**, 1–10 (1982).
53. Lewis, J. W. & Van Essen, D. C. Corticocortical connections of visual, sensorimotor, and multimodal processing areas in the parietal lobe of the macaque monkey. *J. Comp. Neurol.* **428**, 112–137 (2000).
54. Lewis, J. W. & Van Essen, D. C. Mapping of architectonic subdivisions in the macaque monkey, with emphasis on parieto-occipital cortex. *J. Comp. Neurol.* **428**, 79–111 (2000).
55. Van Essen, D. C. *et al.* Mapping visual cortex in monkeys and humans using surface-based atlases. *Vision Res.* **41**, 1359–1378 (2001).
56. Bracewell, R. M., Mazzoni, P., Barash, S. & Andersen, R. A. Motor intention activity in the macaque's lateral intraparietal area. II. Changes of motor plan. *J. Neurophysiol.* **76**, 1457–1464 (1996).

'Rejuvenation' protects neurons in mouse models of Parkinson's disease

C. Savio Chan¹, Jaime N. Guzman¹, Ema Ilijic¹, Jeff N. Mercer¹, Caroline Rick^{1†}, Tatiana Tkatch¹, Gloria E. Meredith² & D. James Surmeier¹

Why dopamine-containing neurons of the brain's substantia nigra pars compacta die in Parkinson's disease has been an enduring mystery. Our studies suggest that the unusual reliance of these neurons on L-type $\text{Ca}_v1.3 \text{ Ca}^{2+}$ channels to drive their maintained, rhythmic pacemaking renders them vulnerable to stressors thought to contribute to disease progression. The reliance on these channels increases with age, as juvenile dopamine-containing neurons in the substantia nigra pars compacta use pacemaking mechanisms common to neurons not affected in Parkinson's disease. These mechanisms remain latent in adulthood, and blocking $\text{Ca}_v1.3 \text{ Ca}^{2+}$ channels in adult neurons induces a reversion to the juvenile form of pacemaking. Such blocking ('rejuvenation') protects these neurons in both *in vitro* and *in vivo* models of Parkinson's disease, pointing to a new strategy that could slow or stop the progression of the disease.

Parkinson's disease (PD) deprives the patient of fluid volitional movement, making common activities like walking or picking up a glass of water difficult or impossible. It is the second most common neurodegenerative disease, rising in incidence after the age of sixty. Although several regions of the brain display signs of pathology¹, the motor symptoms of the disease are unequivocally associated with the degeneration and death of dopamine-containing ('dopaminergic') neurons in the substantia nigra pars compacta (SNc)^{2,3}.

There is no cure for PD, and attempts to slow the progression of the disease have failed, stemming undoubtedly from our poor grasp of disease pathogenesis. Headway on this point has been made by functional studies of genetic mutations found in relatively rare familial forms of PD. Several of the PD genes seem to have roles in mitochondrial function. *PINK1*, *DJ-1*, *OMI/HTRA2* and *LRRK2* encode proteins that are localized at the surface of or in mitochondria^{4–6}. Deficits in mitochondrial complex I function and accumulation of mitochondrial DNA (mtDNA) mutations in the substantia nigra of PD patients add additional weight to the view that mitochondria are important in pathogenesis^{7–9}. As one of the major consequences of mitochondrial generation of reactive oxygen species (ROS), free radicals or peroxynitrites is protein damage, there could be a synergy between mitochondrial and proteosomal dysfunction, hastening cell death¹⁰. However, none of the genes linked to PD is preferentially expressed by SNc dopaminergic neurons, suggesting that their pathogenic impact is modified by some other property that is unique to these neurons.

What is peculiar about SNc dopaminergic neurons, differentiating them from less vulnerable neurons? Early thinking about this question focused on their neurotransmitter, dopamine. There is evidence that oxidation of cytosolic dopamine (and its metabolites) leads to the production of damaging free radicals^{11–13}. However, these studies have relied on manipulations that elevate cytosolic dopamine levels well above normal levels, raising questions about their relevance to normal physiology and PD. There are other reasons to question whether dopamine is a determinant of selective vulnerability. For example, neurons in the ventral tegmental area are not lost at the

same rate in PD as SNc dopaminergic neurons¹⁴. Also, blocking dopamine synthesis does not diminish the sensitivity of dopaminergic neurons to 1-methyl-4-phenyl-1, 2,3,6-tetrahydropyridine (MPTP)¹⁵, a toxin that produces experimental parkinsonism. Lastly, L-DOPA administration (which elevates dopamine levels) in PD patients does not accelerate disease progression and might even be neuroprotective¹⁶, suggesting that at normal cytosolic levels, dopamine is not a significant source of reactive oxidative products and stress.

Adult SNc neurons are Ca^{2+} dependent

If dopamine does not determine vulnerability in PD, then what does? The key to this puzzle might lie in the distinctive physiology of adult SNc dopaminergic neurons. Adult SNc dopaminergic neurons are Ca^{2+} -dependent autonomous pacemakers; that is, their basal activity is intrinsically driven by voltage-dependent L-type Ca^{2+} channels^{17–19}. In tissue slices from the ventral mesencephalon (Fig. 1a, b), this autonomous activity can be seen in cell-attached recordings following blockade of conventional synaptic transmission (Fig. 1c; Supplementary Fig. S1). Bathing cells with the L-type Ca^{2+} channel antagonist isradipine stopped pacemaking ($n = 15$; Fig. 1c). Other members of the dihydropyridine class, like nimodipine, also stopped pacemaking and the underlying membrane potential oscillations that drive the autonomous activity (Fig. 1d). In contrast, blocking voltage-dependent Na^+ channels abolished spikes but left intact the underlying membrane potential oscillation that drives pacemaking (Fig. 1e).

The Ca^{2+} channels that drive pacemaking in SNc dopaminergic neurons are unusual representatives of the L-type channel class. Voltage-dependent Ca^{2+} channels are multimeric proteins in which the pore-forming α -subunit is the principal determinant of gating and pharmacology. L-type channels in brain neurons have one of two α -subunits: $\text{Ca}_v1.2$ or $\text{Ca}_v1.3$. Because of their broad distribution and high affinity for dihydropyridines, channels with the $\text{Ca}_v1.2$ subunit have been the focus of most physiological work. These channels open only at relatively depolarized membrane potentials, making them

¹Department of Physiology, Feinberg School of Medicine, Northwestern University, Chicago, Illinois 60611, USA. ²Department of Cellular and Molecular Pharmacology, Chicago Medical School, Rosalind Franklin University of Medicine and Science, 3333 Green Bay Road, Chicago, Illinois 60064, USA. [†]Present address: BCTU, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK.

poorly suited to the task of driving pacemaking. Channels with pore-forming $\text{Ca}_v1.3$ subunits have a more restricted distribution in the brain, have a low affinity for dihydropyridines and open at relatively hyperpolarized, sub-threshold membrane potentials, making them perfectly suited for pacemaking^{20–23}. SNc dopaminergic neurons robustly expressed $\text{Ca}_v1.3$ subunit messenger RNA (Fig. 1f), suggesting that channels with these subunits, and not those with a $\text{Ca}_v1.2$ subunit, underlie pacemaking in adult SNc dopaminergic neurons.

Surprisingly, unlike acute pharmacological block, genetic deletion of $\text{Ca}_v1.3$ Ca^{2+} channels (encoded by *Cacna1d*) did not stop pacemaking in SNc dopaminergic neurons. In fact, the pacemaking rate in $\text{Ca}_v1.3^{-/-}$ SNc dopaminergic neurons was normal (wild type = 2.5 Hz; $n = 73$; $\text{Ca}_v1.3^{-/-} = 2.4$ Hz, $n = 42$; $P > 0.05$ Mann–Whitney test) and insensitive to isradipine ($n = 6$). Like that in ventral tegmental area dopaminergic neurons, pacemaking in $\text{Ca}_v1.3^{-/-}$ SNc dopaminergic neurons relied on voltage-dependent Na^+ channels to push the membrane potential towards spike threshold (Supplementary Fig. S1). In this respect, SNc $\text{Ca}_v1.3^{-/-}$ dopaminergic neurons also resembled juvenile SNc dopaminergic neurons. In juvenile wild-type SNc dopaminergic neurons, pacemaking was largely unaffected by L-type Ca^{2+} channel antagonists and completely abolished by Na^+ channel antagonists (Fig. 2a), suggesting that adult $\text{Ca}_v1.3^{-/-}$ SNc dopaminergic neurons retain a juvenile form of pacemaking.

The shift in pacemaking mechanism with age seems to be triggered by insertion of $\text{Ca}_v1.3$ Ca^{2+} channels. After the second postnatal week in mice, the contribution of $\text{Ca}_v1.3$ Ca^{2+} channels to somatodendritic L-type Ca^{2+} currents grew, whereas currents attributable to $\text{Ca}_v1.2$ channels remained largely stable (Fig. 2b; Supplementary Fig. S1). This growth in $\text{Ca}_v1.3$ channel currents was not accompanied by any detectable change in Na^+ channel density or gating (data not shown); Na^+ channel density and gating also were not different in $\text{Ca}_v1.3^{-/-}$ SNc dopaminergic neurons (Supplementary Fig. S2). However, the rising density of $\text{Ca}_v1.3$ Ca^{2+} channels was accompanied by the waning importance of another cation channel that contributes to pacemaking. Hyperpolarization-activated and cyclic

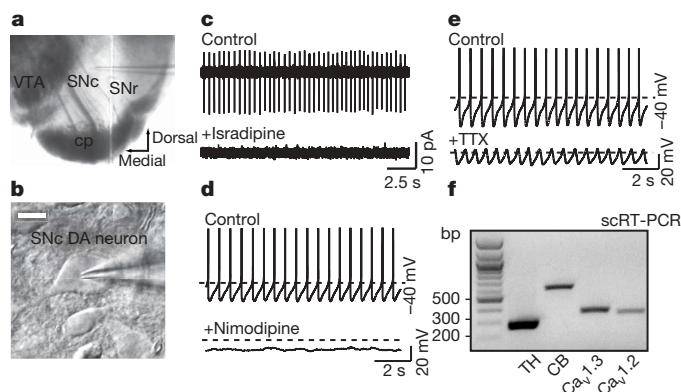


Figure 1 | SNc dopaminergic neurons are $\text{Ca}_v1.3$ - Ca^{2+} -channel-dependent pacemakers. **a**, Light micrograph showing the location of the SNc and adjacent structures in mouse coronal brain slice. VTA, ventral tegmental area, SNr, substantia nigra pars reticulata. **b**, Recordings from SNc dopaminergic neurons were made using IR-DIC video microscopy. Scale bar, 10 μm . **c**, Non-invasive, tight-seal cell-attached patch recordings before and after bath application of Ca_v1 /L-type Ca^{2+} channel antagonist isradipine (20 μM). Note that bathing cells in isradipine stops pacemaking (in 6–23 min, $n = 15$). **d**, Application of a Ca_v1 /L-type Ca^{2+} blocker, nimodipine (10 μM), abolishes autonomous pacemaking in wild-type dopaminergic neurons seen with whole-cell recording ($n = 8$). **e**, In the whole-cell configuration, application of a saturating concentration of TTX (1 μM) abolishes spiking but leaves intact the Ca^{2+} -dependent pacemaker potentials that drive autonomous activity in SNc dopaminergic neurons ($n = 6$). **f**, Single-cell (sc) RT-PCR profile of an SNc dopaminergic neuron showing co-expression of mRNA for TH, calbindin, $\text{Ca}_v1.3$ and $\text{Ca}_v1.2$ ($n = 15$).

nucleotide gated cation (HCN) channels widely partner Na^+ channels in pacemaking²⁴. In $\text{Ca}_v1.3^{-/-}$ SNc dopaminergic neurons, HCN channels were absolutely critical, because an antagonist for these channels (ZD 7288) completely stopped pacemaking (Fig. 2c, middle). In SNc dopaminergic neurons taken from mice less than three weeks old, HCN channels also were very important in maintaining normal spiking rates (Fig. 2c, right). However, as the switch to $\text{Ca}_v1.3$ -channel-dependent pacemaking evolved, these channels became less important (Fig. 2c, left). The waning importance of HCN channels in ventral tier SNc dopaminergic neurons was accompanied by a hyperpolarizing shift in the voltages at which these channels open (Fig. 2d, e; Supplementary Fig. S3), diminishing their engagement during the pacemaking cycle (compare ref. 25) The

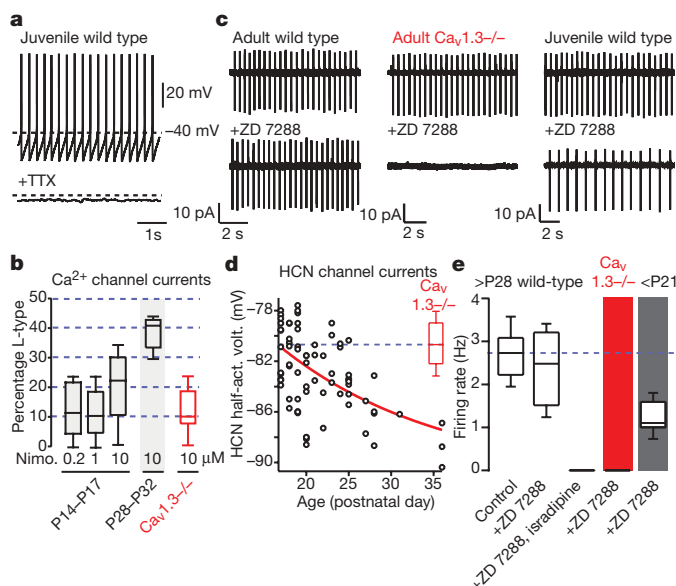


Figure 2 | Pacemaking mechanisms in SNc dopaminergic neurons are developmentally regulated and sensitive to $\text{Ca}_v1.3$ deletion. **a**, Whole-cell recording from a juvenile SNc dopaminergic neuron (P17) revealed the absence of oscillations following TTX (1 μM) block of Na^+ channels. Similar results were seen in other neurons ($n = 10$). **b**, Box plot summarizing the percentage of the total Ca^{2+} current blocked by increasing concentrations of nimodipine in acutely isolated, voltage-clamped juvenile (P14–P17, $n = 6$), adult (P28–P32, $n = 12$) and $\text{Ca}_v1.3^{-/-}$ (red box, P22–P30, $n = 8$) neurons. The magnitude of the block was not increased at higher doses of nimodipine in juvenile neurons ($P > 0.05$, Kruskal–Wallis); however, the block in adult animals was significantly greater than in juvenile neurons ($P < 0.05$, Kruskal–Wallis). **c**, Left, cell-attached recording from an adult (P28) SNc dopaminergic neuron where ZD 7288 had no discernible effect. Centre, similar recording from $\text{Ca}_v1.3^{-/-}$ SNc dopaminergic neuron, showing that ZD 7288 (50 μM) blocked pacemaking. Right, cell-attached patch recording from a juvenile (P18) SNc dopaminergic neuron showing a similar sensitivity to HCN channel blockade. **d**, Plot of half-activation voltage of HCN channels in SNc dopaminergic neurons between the ages of P17 and P36. Note the progressive hyperpolarization of half-activation; the trend was fitted with a simple exponential function. Half-activation voltages for HCN channels in $\text{Ca}_v1.3^{-/-}$ SNc dopaminergic neurons are also plotted. Note the similar voltage dependence seen in SNc dopaminergic neurons from juvenile (<P21) and that from $\text{Ca}_v1.3^{-/-}$ neurons. **e**, Box plot summarizing the change in pacemaking rate with ZD 7288 application. Median discharge rate in adult (>P28), control neurons was near 3 Hz ($n = 73$). Application of ZD 7288 did not significantly reduce the discharge rate in adult neurons ($P > 0.05$, Kruskal–Wallis, $n = 7$, ‘+ZD 7288’), but application of ZD 7288 and isradipine (20 μM) completely blocked activity ($n = 4$, ‘+ZD 7288, isradipine’). In neurons from adult $\text{Ca}_v1.3^{-/-}$ (red highlight, $n = 6$) and juvenile mice (grey highlight, <P21, $n = 7$), pacemaking rate was significantly reduced following ZD 7288 application ($P < 0.05$, Kruskal–Wallis). **b**, **d**, **e**, The edges of the boxes mark the interquartiles of the distribution; the centre line is the median; the ‘whiskers’ mark the extremes of the distribution.

mechanisms underlying this shift remain to be fully explored, but the HCN channels expressed by SNc dopaminergic neurons are exquisitely sensitive to allosteric regulation by cyclic adenosine monophosphate (cAMP)²⁶. Because SNc dopaminergic neurons expressed Ca^{2+} -inhibited isoforms of the enzyme that produces cAMP (adenylyl cyclase types 5, 9) (Supplementary Fig. S4), it is plausible that rising intracellular Ca^{2+} concentrations dampen cAMP synthesis, leading to a shift in HCN gating. Consistent with this hypothesis, suppressing adenylyl cyclase activity normalized the voltage-dependence of HCN channel gating in $\text{Ca}_v1.3^{-/-}$ neurons (Supplementary Fig. S4). Computer simulation of pacemaking also showed that shifting the voltage-dependence of HCN gating was sufficient to re-establish normal pacemaking following deletion of $\text{Ca}_v1.3$ channels (Supplementary Fig. S5). More importantly from a theoretical perspective, the reliance on allosteric regulation of HCN channels to shift the pacemaking mechanism (rather than changes in genetic programming) suggests that the juvenile strategy could be rapidly redeployed if needed.

Rejuvenation of adult SNc neurons

Retention of the ionic mechanisms necessary for the juvenile, Na^+ /HCN-channel-dependent form of pacemaking into adulthood creates an intriguing possibility from the standpoint of PD. It has long been thought that a rising tide of intracellular Ca^{2+} could be a major factor in the death of SNc dopaminergic neurons in PD²⁷. The traditional conjecture has been that late in the disease, depolarization and *N*-methyl-D-aspartate (NMDA) receptor activation triggers excessive Ca^{2+} entry, mitochondrial stress, ROS production and cell death^{28,29}. The increasing reliance with age on $\text{Ca}_v1.3$ Ca^{2+} channels to drive pacemaking suggests that elevations in cytosolic Ca^{2+} might be a major factor in the early stages of the disease, rather than being a late sequela. Rising levels of intracellular Ca^{2+} also can compromise the endoplasmic reticulum and protein processing³⁰. By creating a high level of basal stress, Ca^{2+} loading could synergize with broadly acting environmental toxins or genetic defects, leading to pathogenesis specific to SNc dopaminergic neurons.

How could this potential synergy be disrupted? Pacemaking itself seems to be critical to the function of SNc dopaminergic neurons by maintaining dopamine levels in target structures like the striatum. As a consequence, stopping pacemaking would not be a clinically viable strategy. But what if the Ca^{2+} -dependence of pacemaking could be changed? What if these neurons could be forced to rely on Na^+ /HCN channels? The vast majority of pacemaking neurons in the brain—which do not succumb in PD—rely on these mechanisms, suggesting that they confer some measure of resistance to the factors at play in PD. As shown above, juvenile SNc dopaminergic neurons, which are resistant to toxins that produce a parkinsonian state in adult animals^{31,32}, also rely on Na^+ and HCN channels, rather than $\text{Ca}_v1.3$ Ca^{2+} channels.

The simplest way in which a reversion to juvenile pacemaking might be accomplished is by maintaining the block of $\text{Ca}_v1.3$ channels with a dihydropyridine antagonist, hoping to mimic the situation in the $\text{Ca}_v1.3^{-/-}$ mice. To test this possibility, isradipine (20 μM) was added to the solution bathing brain slices and then the activity of SNc dopaminergic neurons was examined at various times afterward. Thirty minutes after initiating the $\text{Ca}_v1.3$ channel block, SNc dopaminergic neurons were silent (Fig. 3a). However, around an hour after starting channel block, irregular, autonomous spiking was detected in several neurons. This trend continued, so that, after a few hours, almost all of the neurons examined had normal pacemaking, both in terms of rate and regularity (Fig. 3a, b). But now, pacemaking depended not on $\text{Ca}_v1.3$ channels, but Na^+ /HCN channels. Disrupting HCN channels with ZD 7288 completely blocked 'rejuvenated' pacemaking (Fig. 3c). Solidifying the inference that a juvenile form of pacemaking had returned, blocking Na^+ channels not only stopped spiking but also stopped the membrane

potential oscillations underlying pacemaking—just as in juvenile and $\text{Ca}_v1.3^{-/-}$ SNc dopaminergic neurons (Fig. 3d).

Rejuvenation protects SNc neurons

Although the rejuvenation of pacemaking in SNc dopaminergic neurons is a fascinating example of homeostatic plasticity, does it alter the sensitivity to toxins implicated in PD? Systemic administration of the pesticide rotenone leads to the formation of intracellular inclusions resembling Lewy bodies, to the preferential loss of SNc dopaminergic neurons and to behavioural deficits resembling those of PD³³. At the cellular level, the most immediate and profound effects of rotenone are on dendrites, where depolymerization of microtubules leads to beading and fragmentation³⁴. The sensitivity of dendrites probably stems from the fact that Ca^{2+} -dependent pacemaking arises there, creating large oscillations in intracellular Ca^{2+} (ref. 35). These dendritic oscillations seem to be governed largely by Ca^{2+} entry during pacemaking, not by Ca^{2+} entry triggered by Na^+ spikes. Illustrating this point, oscillations in dendritic Ca^{2+} concentration during pacemaking were unaffected by occasional spike failures (Fig. 4a). Moreover, elimination of Na^+ spiking with tetrodotoxin (TTX) had little or no effect on the relative magnitude of the oscillation in dendritic Ca^{2+} concentration (Fig. 4b).

If the impact of rotenone on SNc dopaminergic neurons is dependent on dendritic Ca^{2+} loading during pacemaking, then eliminating the reliance on Ca^{2+} should diminish its effects. As seen previously³⁵, exposure of adult brain slices to rotenone (300 nM) for one hour preferentially stressed the dendrites of SNc dopaminergic neurons, leading to beading and fragmentation (Fig. 5a, centre). Immunoreactivity for tyrosine hydroxylase (TH, a synthesizing enzyme for dopamine and a marker of dopaminergic neurons) also decreased following rotenone exposure. In contrast, dopaminergic neurons in the adjacent ventral tegmental area were virtually unaffected by the same treatment (data not shown). Pre-treatment for two hours with isradipine at a concentration shown to rejuvenate SNc dopaminergic neurons dramatically attenuated the rotenone-induced dendritic loss and fragmentation (Fig. 5a right, c). The

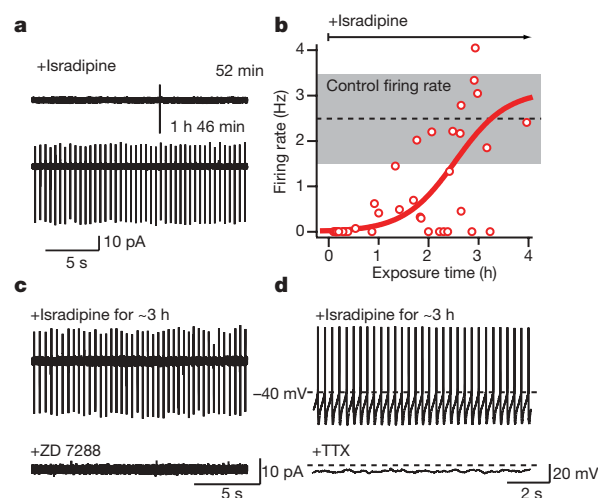


Figure 3 | $\text{Ca}_v1.3$ channel blockade induces a reversion to a juvenile form of Na^+ /HCN-channel-dependent pacemaking. **a**, Short-term blockade of $\text{Ca}_v1.3$ channels with isradipine (20 μM) abolished activity in SNc dopaminergic neurons. Prolonged (~2 h) application of isradipine led to re-establishment of spontaneous activity. **b**, Plot of discharge rate as a function of isradipine incubation time in a sample of SNc dopaminergic neurons; a total of 43 cells were recorded from in all experiments. Smooth line is a fit of the data with a sigmoidal function. **c**, Following re-establishment of pacemaking, application of HCN blocker ZD 7288 (50 μM) disrupted pacemaking in cell-attached recordings ($n = 5$). **d**, TTX (1 μM) also eliminated pacemaking and sub-threshold oscillations, as in juvenile neurons ($n = 6$).

protection against dendritic fragmentation was seen clearly in images generated by confocal laser scanning microscopy where the tissue has been labelled for TH (green) and a dendritic marker, microtubule associated protein 2 (MAP2, red) (Supplementary Fig. S6); again, pretreatment with isradipine led to the retention of normal dendritic morphology of SNc dopaminergic neurons following rotenone exposure (Supplementary Fig. S6). In agreement with the hypothesis that toxicity was dependent on pacemaking-related Ca^{2+} loading, rotenone-induced dendritic fragmentation was not significantly attenuated by TTX but was dramatically diminished in juvenile and $\text{Ca}_v1.3^{-/-}$ SNc dopaminergic neurons (Fig. S6).

To determine whether adult SNc dopaminergic neurons could be rejuvenated *in vivo*, mice were systemically administered isradipine with slow-release, subcutaneous pellets. SNc dopaminergic neurons in brain slices prepared from mice treated for 7 days were virtually insensitive to isradipine, in sharp contrast to age-matched, naive neurons (Fig. 5d), demonstrating that the switch in pacemaking could be accomplished *in vivo*. As previously reported for another dihydropyridine antagonist, isradipine treatment failed to protect SNc dopaminergic terminals and cell bodies against large intrastriatal injections of the toxin 6-hydroxydopamine (6-OHDA). However, isradipine treatment provided a robust protection against a more modest intrastriatal 6-OHDA challenge (Supplementary Fig. S7). Another commonly used experimental model of PD in mice and primates relies on systemic injection of the toxin MPTP^{31,36}. MPTP crosses the blood–brain barrier and is converted to 1-methyl-4-phenylpyridinium (MPP⁺); MPP⁺ is taken up by dopaminergic neurons through a plasma membrane dopamine transporter and inhibits mitochondrial respiratory metabolism, promoting cell death. To determine whether

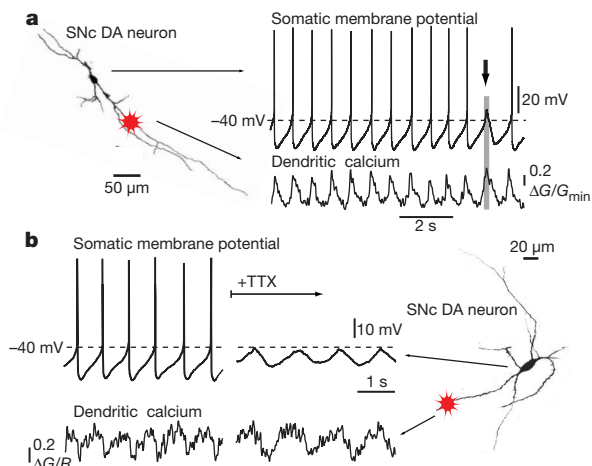


Figure 4 | Pacemaking currents govern dendritic Ca^{2+} concentrations. **a**, Left, an image of an SNc dopaminergic (DA) neuron (P29) obtained with two photon laser scanning microscopy; neurons were filled with Alexa594 (50 μM) and Fluo-4 (100 μM) through whole-cell recording glass pipette; calcium imaging was performed by line scanning along dendritic segment (location is indicated by red star) $\sim 100 \mu\text{m}$ from the soma. Right, plot of somatic voltage (measured with the somatic electrode) and co-incident dendritic Ca^{2+} signal during normal pacemaking (marked by the red star). G, green fluorescence, ΔG , the change in the green fluorescence, G_{\min} , minimal green fluorescence. Arrow marks a time point at which the somatic spike failed but the dendritic Ca^{2+} oscillation occurred normally, indicating that the two events were largely independent. Similar results were seen in other neurons examined ($n = 3$). Voltage-clamping the soma at -60 mV to eliminate spiking produced similar results (data not shown, $n = 3$). **b**, Left, Somatic recordings of membrane potential in SNc dopaminergic neuron (P26) before and after application of TTX to block spike generation. Below are the Ca^{2+} signals generated in a dendritic region (marked with the star); note that the Ca^{2+} oscillations were largely unaffected by TTX. Similar results were seen in all neurons tested ($n = 4$). R, red fluorescence. Right, an image of the Alexa594-filled SNc dopaminergic neuron recorded from, as above.

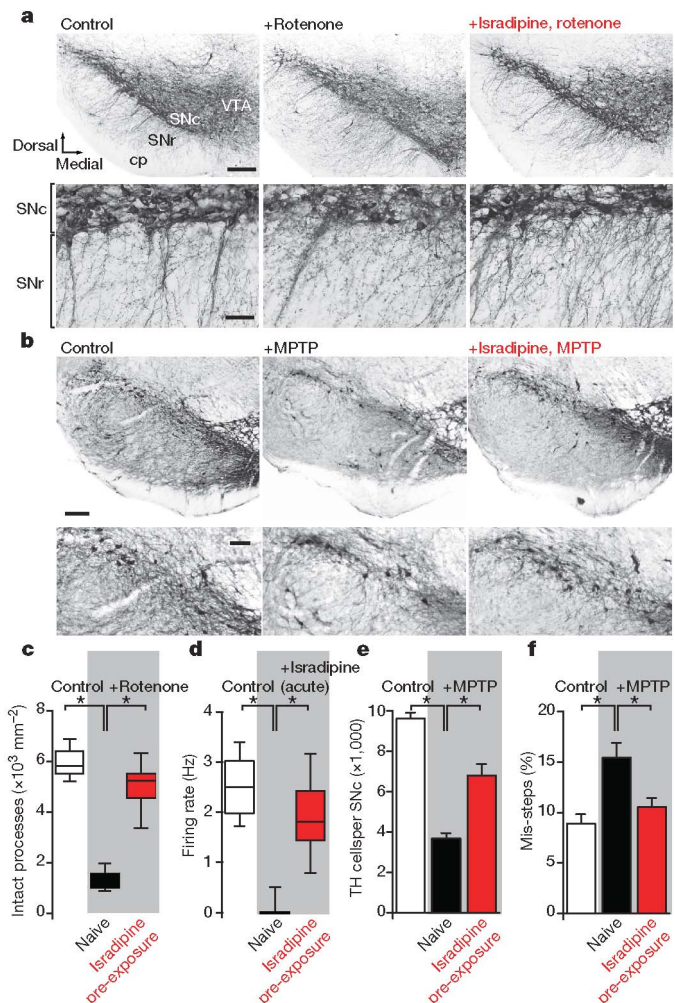


Figure 5 | Rejuvenation of SNc dopaminergic neurons protects them against the mitochondrial toxin rotenone and chronic MPTP treatment. **a**, Photomicrographs of brain slices from mesencephalon held in saline (left) or rotenone (300 nM) for 1 h (centre, right); rotenone-treated slices were pre-incubated with saline (centre) or isradipine (20 μM) (right) for 2 h. Slices were fixed and processed for tyrosine hydroxylase immunoreactivity. Top, low magnification, scale bar, 200 μm ; bottom, high magnification, scale bar, 50 μm . Note the sparing of TH⁺ processes running through the SNr following isradipine incubation. **b**, Light micrographs of sections processed for TH immunoreactivity; sections taken from control (left), MPTP (centre) and isradipine + MPTP (right)-treated mice. Top, low magnification, scale bar, 100 μm ; bottom, high magnification, scale bar, 50 μm . Note the sparing of TH⁺ neurons throughout the lateromedial extent of SNc following systemic, extended-release isradipine treatment. **c**, Box plot summarizing the protection against dendritic fragmentation produced by isradipine treatment; control (open box, $n = 9$); rotenone (filled box, $n = 9$); isradipine + rotenone (red box, $n = 9$); (* $P < 0.001$, Kruskal–Wallis). Note that in box plots, the central line is the median value, the box edges mark the inter-quartiles and the bars mark the limits of the sample distribution. **d**, Box plot summarizing the effects of acute application of isradipine (20 μM) on pacemaking in dopaminergic neurons from naive/placebo animals (filled box, $n = 20$) and those treated with systemically administered isradipine (red box, $n = 13$). For comparison, a box plot of the discharge rate in untreated control neurons is shown at the left (open box, $n = 73$). **e**, Stereological analysis of SNc dopaminergic cell loss following MPTP treatment. Isradipine administration attenuated cell loss induced by MPTP (* $P < 0.05$, one-way ANOVA). **f**, Behavioural assessment revealed no significant motor impairment in MPTP-treated animals pre-exposed to systemic isradipine administration (red, * $P < 0.05$, one-way ANOVA, error bars show the standard error of the mean). Data derived from isradipine treated animals are highlighted in red.

'rejuvenation' would afford protection in this model, isradipine-treated adult mice were subjected to repeated injections of the toxin MPTP over the course of 5 weeks. This chronic MPTP model of PD produces a substantial loss of SNc dopaminergic neurons and motor deficits³⁷. On the basis of stereological analysis of TH-immunoreactive and Nissl-stained sections, isradipine treatment reduced the MPTP-induced loss of SNc dopaminergic neurons by nearly half (Fig. 5b, e) and prevented the development of motor deficits (Fig. 5f). Protection was not attributable to altered metabolism, as brain MPP⁺ levels following an MPTP challenge were not affected by isradipine treatment (control = 103.5 ng per mg wet tissue, $n = 9$; isradipine = 102.5 ng per mg wet tissue, $n = 7$; $P > 0.05$, Mann–Whitney rank sum test). Moreover, preliminary experiments suggest that genetic deletion of Ca_v1.3 channels provides SNc dopaminergic neurons the same protection against MPTP toxicity afforded by isradipine treatment ($n = 3$).

Taken together, these studies point to a theory of pathogenesis in PD that explains the differential vulnerability of SNc dopaminergic neurons. What sets these neurons apart is their reliance on voltage-dependent Ca²⁺ channels to drive autonomous pacemaking. This reliance increases with age and leads to sustained elevations in cytosolic Ca²⁺ concentration, particularly in dendrites³⁸. Cytosolic Ca²⁺ stimulates mitochondrial respiratory metabolism and ROS generation—long hypothesized to be key steps in dopaminergic cell death^{10,31,39}. This effect is mediated in part by the cellular ATP demand posed by the need to sequester or pump Ca²⁺ out of cells, but it also appears to reflect the role of mitochondria in intracellular Ca²⁺ dynamics. Recent work has shown that a polymorphism in mtDNA associated with increased risk of PD compromises the ability of mitochondria to buffer cytosolic Ca²⁺ transients⁴⁰. Along these same lines, the expression of mobile Ca²⁺-binding proteins seems to confer some measure of neuronal protection in PD⁴¹. ROS generated by stimulation of mitochondrial respiration have been linked to high mtDNA mutation rates, compromised complex I function and accumulation of mis-folded or damaged proteins in Lewy bodies^{7,8,10}. This stress should be particularly prominent in SNc dopaminergic neuronal dendrites, helping to account for their retraction before somatic atrophy with ageing⁴². The accumulation of defects with age could make lethal normally sustainable secondary insults or 'hits' to mitochondrial function, arising either from broadly acting environmental toxins or somatic genetic mutations.

The contention that age and activity-dependent Ca²⁺ influx is a central determinant of PD pathogenesis is consistent with a theory of PD staging recently advanced by Braak¹. Braak contends that Lewy body inclusions are the hallmark of PD and that the formation of these inclusions extends well beyond the SNc, actually beginning in the caudal medulla. He speculates that the feature linking vulnerable neurons is a long, thin, unmyelinated axon that imposes a metabolic load. The metabolic demands associated with an axon of this type are closely linked to Ca²⁺ homeostasis⁴³. In fine axons and terminals, where the surface area to volume ratio is high, entry through voltage-dependent channels can significantly elevate intracellular Ca²⁺ concentrations, stressing mitochondria and promoting degenerative changes⁴³. Although attractive in several respects, key points of Braak's hypothesis are controversial and remain to be tested. For example, there is agreement that the appearance of Lewy bodies is a sign of cellular stress, but there is not agreement that they are a sign of imminent cell death or even cellular dysfunction. If axonal stress is critical to Lewy body formation, the presence of a concomitant dendritic Ca²⁺-dependent stress in SNc dopaminergic neurons could explain why they are among the first neurons to die in PD, even though they form Lewy bodies relatively late in the progression of the disease.

If our hypothesis about PD pathogenesis is correct, then there is a way of preventing or slowing the progression of the disease, at least in SNc dopaminergic neurons. Blocking Ca_v1.3 L-type Ca²⁺ channels that drive adult pacemaking induces a homeostatic reversion of

SNc dopaminergic neurons to the Ca²⁺-independent pacemaking mechanisms used early in life. 'Rejuvenation' conferred a significant reduction in sensitivity to toxins used to create experimental PD, in agreement with previous studies showing they have relatively little effect on juvenile SNc dopaminergic neurons but potentially kill adult neurons^{32,36}. This conclusion also is consistent with previous studies examining the impact of a less effective Ca_v1.3 channel antagonist on MPTP toxicity in rodents and primates^{44,45}. The fact that the 'rejuvenation' can be brought about by treatment with the dihydropyridine isradipine—a very well-tolerated drug that is widely used in the treatment of hypertension and stroke^{46,47}—points to a neuroprotective strategy that could be tried immediately. A recent retrospective study of hypertensive patients suggests that dihydropyridine use might lower the incidence of PD by 30–50% (ref. 48). Diminishing the vulnerability of SNc dopaminergic neurons should not only decrease the incidence of PD but also slow its progression, broadening the therapeutic window for PD patients in the early stages of the disease.

METHODS SUMMARY

Male C57BL/6 mice were used. For the purposes of our study, juvenile animals were defined as less than 21 days postnatal (P21); adult animals were defined as older than P28. Mice with a null mutation of the Ca_v1.3 gene were obtained from J. Striessnig and re-derived. Conventional tight-seal whole-cell patch-clamp and cell-attached recordings were made on visually identified dopaminergic neurons in mesencephalon slices with procedures similar to those previously described⁴⁹. Procedures for recording total Ca²⁺ currents have been previously described²¹. For dendritic fluorescence measurements a Ti:sapphire pulsed laser was tuned to 810 nm. Fluorescence emission was collected during line scans along dendritic segments in epifluorescence mode by external or non-descanned photomultiplier tubes. Primers and reaction protocols for single-cell RT–PCR have been described previously²¹. For systemic isradipine administration mice were implanted with biodegradable-carrier bound isradipine (3 µg g⁻¹ d⁻¹) pellets or inert placebo control pellets. For rotenone (300 nM) toxicity assay, brain slices were incubated in ACSF for 2 h with or without isradipine (20 µM) followed by incubation with rotenone for 1 h and subjected to immunohistochemical analysis. Prior to MPTP lesioning, mice were implanted with isradipine or placebo pellets; one week later, animals were administered MPTP using a chronic treatment protocol; at the end of treatment, animals were subjected to a grid test, and processed as described elsewhere³⁷. MPTP metabolism to MPP⁺ was assayed with extraction and analytical procedures similar to those previously published⁵⁰. Striatal 6-OHDA (1 µg per 1 µl) lesions were performed with mice implanted with isradipine or placebo pellets. Lesions were performed 7 days after pellet implantation. Immunocytochemical analysis was performed 3 weeks after lesioning. Small samples ($n < 10$) were analysed with non-parametric Kruskal–Wallis analysis of variance (ANOVA); larger samples were analysed with a parametric one-way ANOVA followed by a pairwise multiple comparison procedure (Student–Newman–Keuls method).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 20 September 2006; accepted 16 April 2007.

Published online 10 June 2007.

1. Braak, H., Ghebremedhin, E., Rub, U., Bratzke, H. & Del Tredici, K. Stages in the development of Parkinson's disease-related pathology. *Cell Tissue Res.* **318**, 121–134 (2004).
2. Hornykiewicz, O. Dopamine (3-hydroxytyramine) and brain function. *Pharmacol. Rev.* **18**, 925–964 (1966).
3. Riederer, P. & Wuketich, S. Time course of nigrostriatal degeneration in Parkinson's disease. A detailed study of influential factors in human brain amine analysis. *J. Neural Transm.* **38**, 277–301 (1976).
4. Abou-Sleiman, P. M., Muqit, M. M. & Wood, N. W. Expanding insights of mitochondrial dysfunction in Parkinson's disease. *Nature Rev. Neurosci.* **7**, 207–219 (2006).
5. Zhang, L. et al. Mitochondrial localization of the Parkinson's disease related protein DJ-1: Implications for pathogenesis. *Hum. Mol. Genet.* **14**, 2063–2073 (2005).
6. Kwong, J. Q., Beal, M. F. & Manfredi, G. The role of mitochondria in inherited neurodegenerative diseases. *J. Neurochem.* **97**, 1659–1675 (2006).
7. Bender, A. et al. High levels of mitochondrial DNA deletions in substantia nigra neurons in aging and Parkinson disease. *Nature Genet.* **38**, 515–517 (2006).

8. Kraytsberg, Y. *et al.* Mitochondrial DNA deletions are abundant and cause functional impairment in aged human substantia nigra neurons. *Nature Genet.* **38**, 518–520 (2006).
9. Swerdlow, R. H. *et al.* Origin and functional consequences of the complex I defect in Parkinson's disease. *Ann. Neurol.* **40**, 663–671 (1996).
10. Dawson, T. M. & Dawson, V. L. Molecular pathways of neurodegeneration in Parkinson's disease. *Science* **302**, 819–822 (2003).
11. Greenamyre, J. T. & Hastings, T. G. Biomedicine. Parkinson's—divergent causes, convergent mechanisms. *Science* **304**, 1120–1122 (2004).
12. Zecca, L., Zucca, F. A., Wilms, H. & Sulzer, D. Neuromelanin of the substantia nigra: A neuronal black hole with protective and toxic characteristics. *Trends Neurosci.* **26**, 578–580 (2003).
13. Michel, P. P. & Hefti, F. Toxicity of 6-hydroxydopamine and dopamine for dopaminergic neurons in culture. *J. Neurosci. Res.* **26**, 428–435 (1990).
14. Kish, S. J., Shannak, K. & Hornykiewicz, O. Uneven pattern of dopamine loss in the striatum of patients with idiopathic Parkinson's disease. Pathophysiologic and clinical implications. *N. Engl. J. Med.* **318**, 876–880 (1988).
15. Hasbani, D. M., Perez, F. A., Palmiter, R. D. & O'Malley, K. L. Dopamine depletion does not protect against acute 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine toxicity *in vivo*. *J. Neurosci.* **25**, 9428–9433 (2005).
16. Fahn, S. Does levodopa slow or hasten the rate of progression of Parkinson's disease? *J. Neurol.* **252** (Suppl. 4), IV37–IV42 (2005).
17. Mercuri, N. B. *et al.* Effects of dihydropyridine calcium antagonists on rat midbrain dopaminergic neurones. *Br. J. Pharmacol.* **113**, 831–838 (1994).
18. Nedergaard, S., Flatman, J. A. & Engberg, I. Nifedipine- and omega-conotoxin-sensitive Ca^{2+} conductances in guinea-pig substantia nigra pars compacta neurones. *J. Physiol. (Lond.)* **466**, 727–747 (1993).
19. Shepard, P. D. & Stump, D. Nifedipine blocks apamin-induced bursting activity in nigral dopamine-containing neurons. *Brain Res.* **817**, 104–109 (1999).
20. Xu, W. & Lipscombe, D. Neuronal $\text{Ca}_v1.3\alpha_1$ L-type channels activate at relatively hyperpolarized membrane potentials and are incompletely inhibited by dihydropyridines. *J. Neurosci.* **21**, 5944–5951 (2001).
21. Olson, P. A. *et al.* G-protein-coupled receptor modulation of striatal $\text{Ca}_v1.3$ L-type Ca^{2+} channels is dependent on a Shank-binding domain. *J. Neurosci.* **25**, 1050–1062 (2005).
22. Koschak, A. *et al.* $\alpha_1\text{D}$ ($\text{Ca}_v1.3$) subunits can form L-type Ca^{2+} channels activating at negative voltages. *J. Biol. Chem.* **276**, 22100–22106 (2001).
23. Scholze, A., Plant, T. D., Dolphin, A. C. & Nurnberg, B. Functional expression and characterization of a voltage-gated $\text{Ca}_v1.3$ ($\alpha_{1\text{D}}$) calcium channel subunit from an insulin-secreting cell line. *Mol. Endocrinol.* **15**, 1211–1221 (2001).
24. Surmeier, D. J., Mercer, J. N. & Chan, C. S. Autonomous pacemakers in the basal ganglia: Who needs excitatory synapses anyway? *Curr. Opin. Neurobiol.* **15**, 312–318 (2005).
25. Zolles, G. *et al.* Pacemaking by HCN channels requires interaction with phosphoinositides. *Neuron* **52**, 1027–1036 (2006).
26. Neuhoff, H., Neu, A., Liss, B. & Roeper, J. I_h channels contribute to the different functional properties of identified dopaminergic subpopulations in the midbrain. *J. Neurosci.* **22**, 1290–1302 (2002).
27. Beal, M. F. Excitotoxicity and nitric oxide in Parkinson's disease pathogenesis. *Ann. Neurol.* **44**, S110–S114 (1998).
28. Coyle, J. T. & Puttfarcken, P. Oxidative stress, glutamate, and neurodegenerative disorders. *Science* **262**, 689–695 (1993).
29. Krieger, C. & Duchen, M. R. Mitochondria, Ca^{2+} and neurodegenerative disease. *Eur. J. Pharmacol.* **447**, 177–188 (2002).
30. Verkhratsky, A. Physiology and pathophysiology of the calcium store in the endoplasmic reticulum of neurons. *Physiol. Rev.* **85**, 201–279 (2005).
31. Dauer, W. & Przedborski, S. Parkinson's disease: mechanisms and models. *Neuron* **39**, 889–909 (2003).
32. Phinney, A. L. *et al.* Enhanced sensitivity of dopaminergic neurons to rotenone-induced toxicity with aging. *Parkinsonism Relat. Disord.* **12**, 228–238 (2006).
33. Betarbet, R. *et al.* Chronic systemic pesticide exposure reproduces features of Parkinson's disease. *Nature Neurosci.* **3**, 1301–1306 (2000).
34. Jiang, Q., Yan, Z. & Feng, J. Activation of group III metabotropic glutamate receptors attenuates rotenone toxicity on dopaminergic neurons through a microtubule-dependent mechanism. *J. Neurosci.* **26**, 4318–4328 (2006).
35. Bywood, P. T. & Johnson, S. M. Mitochondrial complex inhibitors preferentially damage substantia nigra dopamine neurons in rat brain slices. *Exp. Neurol.* **179**, 47–59 (2003).
36. Przedborski, S. & Vila, M. The 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine mouse model: A tool to explore the pathogenesis of Parkinson's disease. *Ann. NY Acad. Sci.* **991**, 189–198 (2003).
37. Petroske, E., Meredith, G. E., Callen, S., Totterdell, S. & Lau, Y. S. Mouse model of Parkinsonism: A comparison between subacute MPTP and chronic MPTP/probenecid treatment. *Neuroscience* **106**, 589–601 (2001).
38. Wilson, C. J. & Callaway, J. C. Coupled oscillator model of the dopaminergic neuron of the substantia nigra. *J. Neurophysiol.* **83**, 3084–3100 (2000).
39. Orrenius, S., Zhivotovsky, B. & Nicotera, P. Regulation of cell death: The calcium-apoptosis link. *Nature Rev. Mol. Cell Biol.* **4**, 552–565 (2003).
40. Kazuno, A. A. *et al.* Identification of mitochondrial DNA polymorphisms that alter mitochondrial matrix pH and intracellular calcium dynamics. *PLoS Genet.* **2**, e128 (2006).
41. Yamada, T., McGeer, P. L., Baimbridge, K. G. & McGeer, E. G. Relative sparing in Parkinson's disease of substantia nigra dopamine neurons containing calbindin-D28K. *Brain Res.* **526**, 303–307 (1990).
42. Ai, Y. *et al.* Intraputaminal infusion of GDNF in aged rhesus monkeys: Distribution and dopaminergic effects. *J. Comp. Neurol.* **461**, 250–261 (2003).
43. Coleman, M. Axon degeneration mechanisms: Commonality amid diversity. *Nature Rev. Neurosci.* **6**, 889–898 (2005).
44. Kupsch, A. *et al.* Pretreatment with nimodipine prevents MPTP-induced neurotoxicity at the nigral, but not at the striatal level in mice. *Neuroreport* **6**, 621–625 (1995).
45. Kupsch, A. *et al.* 1-Methyl-4-phenyl-1,2,3,6-tetrahydropyridine-induced neurotoxicity in non-human primates is antagonized by pretreatment with nimodipine at the nigral, but not at the striatal level. *Brain Res.* **741**, 185–196 (1996).
46. Grossman, E., Messerli, F. H., Oren, S., Nunez, B. & Garavaglia, G. E. Cardiovascular effects of isradipine in essential hypertension. *Am. J. Cardiol.* **68**, 65–70 (1991).
47. Johnson, B. A., Devous, M. D. Sr, Ruiz, P. & Ait-Daoud, N. Treatment advances for cocaine-induced ischemic stroke: Focus on dihydropyridine-class calcium channel antagonists. *Am. J. Psychiatry* **158**, 1191–1198 (2001).
48. Rodnitzky, R. L. Can calcium antagonists provide a neuroprotective effect in Parkinson's disease? *Drugs* **57**, 845–849 (1999).
49. Chan, C. S., Shigemoto, R., Mercer, J. N. & Surmeier, D. J. HCN2 and HCN1 channels govern the regularity of autonomous pacemaking and synaptic resetting in globus pallidus neurons. *J. Neurosci.* **24**, 9921–9932 (2004).
50. Hows, M. E. *et al.* High-performance liquid chromatography/tandem mass spectrometry assay for the determination of 1-methyl-4-phenyl pyridinium (MPP⁺) in brain tissue homogenates. *J. Neurosci. Methods* **137**, 221–226 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by grants from the Picower Foundation and NIH NINDS to D.J.S. and G.E.M. We thank J. Held, D. Wokosin, P. Hockberger, E. Mugnaini, S. Ulrich, K. Saporito, Q. Ruan, J. Jackolin, F. Jodelka and M. Avram for help with this work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.J.S. (j-surmeier@northwestern.edu).

METHODS

Animals. Male C57BL/6 mice (Charles River) were used. For the purposes of our study, juvenile animals were defined as less than 21 days postnatal (P21); adult animals were defined as older than P28. Mice with a null mutation of the $Ca_v1.3$ gene were obtained from J. Striessnig (Institut für Biochemische Pharmakologie, Innsbruck, Austria)⁵¹, re-derived, and backcrossed on a C57BL/6 background in the Northwestern University facility. Experiments with congenic wild-type and $Ca_v1.3^{-/-}$ littermates were used on the same day whenever possible. The handling of mice and all procedures performed on them were approved by the institutional Animal Care and Use Committee and were in accordance with the 'National Institutes of Health Guide to the Care and Use of Laboratory Animals' and Society for Neuroscience guidelines.

Isradipine treatment. To test for reversion of the pacemaking mechanism, mice (P25–P28) were anesthetized with a mixture of ketamine and xylazine; extended release pellets with biodegradable-carrier bound isradipine (21 d release, $3 \mu\text{g g}^{-1} \text{d}^{-1}$) or inert placebo control pellets (Innovative Research of America) were implanted into the interscapular space. Two to four animals were housed per cage with food and water *ad libitum*. The room was maintained at constant temperature on a 12-h light–dark cycle. One week later, animals were killed, brain removed and mesencephalon slices prepared in chilled saline as described above. Slices were allowed to recover briefly (~ 30 min at 34°C) in an artificial CSF (ACSF) containing isradipine ($20 \mu\text{M}$). Pacemaking was assessed with cell-attached recordings (in the presence of isradipine) in the next 30 min.

MPTP treatment. Eight- to ten-week-old mice, weighing 21–24 g at the beginning of the study, were used. Animals were either implanted with isradipine or placebo time-release pellets (60 d release, $3 \mu\text{g g}^{-1} \text{d}^{-1}$) as described above. Mice were administered MPTP following a chronic published protocol³⁷. Briefly, mice were treated with 10 doses of MPTP hydrochloride (25 mg kg^{-1} in saline, s.c.) and an adjuvant, probenecid (250 mg kg^{-1} in Tris-HCl buffer, i.p.), on a five-week schedule with an interval of 3.5 d between consecutive doses. Controls were treated with vehicle (saline) in the same volume. Probenecid was used to inhibit the rapid clearance of MPTP. Earlier work has shown no difference in the number of dopaminergic neurons or in the concentration of striatal dopamine after probenecid or saline treatments³⁷.

Terminal field 6-OHDA lesions. Male C57BL/6 mice at P28–P31 were implanted with isradipine (21 d release, $3 \mu\text{g g}^{-1} \text{d}^{-1}$) or placebo time-release pellets. Lesions were performed 7 d later. In brief, animals were anesthetized with a mixture of ketamine (50 mg kg^{-1} i.p.) and xylazine (4.5 mg kg^{-1} i.p.). After immobilization on a stereotaxic frame (model 940, David Kopf Instruments) with a Cunningham adaptor (Harvard Apparatus), a hole was drilled (~ 1 mm diameter) at 0.4 mm posterior and 1.6 mm lateral to bregma for injection into the caudate putamen. 6-OHDA (Sigma) was dissolved at a concentration of $1 \mu\text{g}$ per μl saline with 0.2 mg ml^{-1} ascorbic acid and injected in final dosages of $1 \mu\text{g}$. Injection of 6-OHDA ($1 \mu\text{l}$) was performed using a calibrated glass micropipette (2-000-00, Drummond Scientific Company), with its tip at a depth of 3.2 mm from dura; 6-OHDA was ejected at a rate of $0.05 \mu\text{l min}^{-1}$. The micropipette was left *in situ* for another 60 min after the injection to maximize tissue retention of 6-OHDA. Immunohistochemical analyses were performed 3 weeks later.

Tissue preparation. Animals were anesthetized with isoflurane and decapitated. Brains were removed rapidly and placed immediately in ice-cold ACSF containing (in mM): 125 NaCl, 2.5 KCl, 1 MgCl_2 , 2 CaCl_2 , 1.25 NaH_2PO_4 , 25 glucose, and 25 NaHCO_3 , bubbled continuously with 'carbogen' (95% O_2 and 5% CO_2). Thin coronal mesencephalon slices (220 – $275 \mu\text{m}$) containing SNc were made using a vibrating microtome (VT1000 s, Leica Instrument) and equilibrated in ACSF or a low- Ca^{2+} buffer containing (in mM) 140 Na isethionate, 23 glucose, 15 HEPES, 2 KCl, 4 MgCl_2 , 0.2 CaCl_2 , 1 kynurenic acid, 0.1 *N*-nitro-L-arginine, and 0.005 glutathione, pH 7.4, 300 – 305 mOsm l^{-1} before being incubated for 1–5 h in sodium bicarbonate-buffered Earle's balanced salt solution (EBSS) bubbled with 'carbogen'. EBSS also contained the following (in mM): 23 glucose, 1 kynurenic acid, 0.1 *N*-nitro-L-arginine, and 0.005 glutathione.

Whole-cell and cell-attached recording in slices. Slices were transferred to a small volume (<0.5 ml) recording chamber that was mounted on a fixed-stage, upright microscope (BX51, Olympus America) equipped with IR-DIC (0.9 numerical aperture, NA) with de Sénarmont compensation (Olympus). Experiments were performed at $34 \pm 1^\circ\text{C}$ unless specified otherwise. The recording chamber was superfused with carbogen-saturated ACSF with a flow rate of 2 – 3 ml min^{-1} running through an in-line heater (SH-27B with TC-324B controller; Warner Instruments). Neuronal somata and proximal dendrites were visualized by video microscopy at high magnification ($60\times$, 0.9 NA water immersion objective, Olympus) with a back-thinned, frame-transfer cooled-CCD camera (Micromax EBFT512; Roper Scientific) aided by a contrast enhancement system (Argus-20, Hamamatsu Photonics).

Only neurons in the mid to ventral tier of SNc were studied because of their uniform, well documented electrophysiological properties as well as the reliance of their autonomous firing on the dihydropyridine-sensitive Ca_v1/L -type Ca^{2+} channels. Following previous studies, the SNc was defined as the region between the medial lemniscus dorsally and the substantia nigra pars reticulata ventrally^{52,53}. This region contains a dense band of neurons with large cell bodies and with dendritic processes running in a mediolateral manner when visualized under IR-DIC microscopy. Conventional tight-seal ($>3 \text{ G}\Omega$) whole-cell patch-clamp and cell-attached recordings were made on visually identified, dopaminergic neurons, based on size and somatodendritic morphology. Dopaminergic neurons were identified by their physiological features^{54–56}. These included resting level of discharge (~ 1 – 4 Hz during cell-attached recording), a prominent voltage-sag during hyperpolarizing current injection and high spike threshold (typically above -40 mV). Identification of neurons was verified by single cell RT-PCR for tyrosine hydroxylase (dopaminergic neuron) and glutamate decarboxylase (non-dopaminergic neuron).

Patch electrodes (1.5 mm outer diameter) were fabricated from filamented, thick-wall borosilicate glass (Sutter Instruments) pulled on a Flaming–Brown puller (P-97, Sutter) and fire-polished immediately before use. Pipette resistance was typically ~ 3 – $6 \text{ M}\Omega$ when filled with recording solution. The recording internal solution consisted of (in mM): 140 KMeSO_4 , 5 KCl, 10 Na-phosphocreatine, 0.025–0.05 EGTA, 2.0 Mg-ATP , 0.4 $\text{Na}_3\text{-GTP}$, 10 HEPES, pH 7.25–7.30, 280 mOsm l^{-1} . EGTA was omitted from the solution for Ca^{2+} imaging experiments. For identification, $50 \mu\text{M}$ Alexa Fluor 594 (Invitrogen) or 0.2% biocytin were added in a subset of experiments. The liquid junction potential in our recording ACSF was $\sim 7 \text{ mV}$ and not corrected for. Somatic whole-cell patch-clamp recordings were obtained via a MultiClamp 700B amplifier (Molecular Devices) interfaced to a Pentium-based PC running pClamp9 (Molecular Devices). The signal was filtered at 1–4 kHz and digitized at 5–20 kHz with a Digidata 1322A (Molecular Devices). For current-clamp recordings, the amplifier bridge circuit was adjusted to compensate for electrode resistance and monitored. Electrode capacitance was also compensated. If series resistance increased $>20\%$ during recording, the data were discarded.

Two-photon Ca^{2+} imaging. SNc dopaminergic neurons were filled with Alexa Fluor 594 ($50 \mu\text{M}$) and Fluo-4 ($100 \mu\text{M}$; Invitrogen) through the patch pipette. Dyes were allowed to equilibrate for at least 15 min before imaging. Dendritic fluorescence measurements used a Ti:sapphire pulsed laser (90-MHz; Chameleon-XR, Coherent) tuned to 810 nm for simultaneous excitation of 'red' (Alexa Fluor 594) and 'green' (Fluo-4) dyes. Laser-scanned images were acquired with a Bio-Rad Radiance MPD system. Fluorescence emission was collected during line scans (500 – 600 Hz) along dendritic segments in epifluorescence mode by external or non-descanned photomultiplier tubes (PMTs). The green fluorescence (500 – 550 nm) was detected by a bialkali-cathode PMT, and the red fluorescence (570 – 620 nm) was collected by a multialkali-cathode (S-20) PMT. Image acquisition, electrophysiological acquisition and off-line data analyses were performed with shareware (WinFluor and PicViewer) written by J. Dempster, Strathclyde University, Glasgow, UK.

Whole-cell voltage-clamp recording of acutely isolated neurons. Whole-cell recordings were performed using electrodes pulled from Corning 7052 glass (1.2 mm outer diameter), coated with R-6101 (Corning), and fire-polished immediately before use. Electrodes typically had resistances of 1.5 – $2.5 \text{ M}\Omega$ in the bath. Recordings were obtained at room temperature (20 – 22°C) with an Axopatch 200B amplifier (Molecular Devices) interfaced to a Macintosh G4 computer (Apple Computers) running Pulse software (HEKA Elektronik). After the gigohm seal was formed and the cell membrane ruptured, series resistance was compensated (75 – 80%) and frequently monitored. During recording, cells were bathed in extracellular solutions applied via a gravity-fed capillary perfusion array positioned several hundred micrometres away from the cell under study. Solutions were changed by adjusting the position of the array using a DC motor (Newport). Solution changes were complete within $<1 \text{ s}$.

Procedures for recording rapidly inactivating Na^+ currents have previously been described³⁷. In brief, the external solution contained the following (in mM): 10 NaCl, 110 tetraethylammonium (TEA) chloride, 10 CsCl, 10 HEPES, 2 BaCl_2 , 1 MgCl_2 , and 0.3 CdCl_2 , pH 7.4 (300 – 305 mOsm l^{-1}). For recording persistent and resurgent Na^+ currents, the external solution contained the following (in mM): 115 NaCl, 45 TEA chloride, and 2 BaCl_2 , 1 MgCl_2 , 0.3 CdCl_2 , and 10 HEPES, pH 7.4 (300 – 305 mOsm l^{-1}). The liquid junction potential ($<2 \text{ mV}$) was not compensated for. The intracellular recording solution contained the following (in mM): 70 *N*-methyl-D-glucamine, 20 HEPES, 50 Cs_2SO_4 , 2 MgCl_2 , 0.5 Na_2SO_4 , 12 phosphocreatine, 2 Mg-ATP , 0.7 $\text{Na}_2\text{-GTP}$, and 0.1 leupeptin, pH 7.25, with H_2SO_4 (265 – 270 mOsm l^{-1}). Procedures for recording total Ca^{2+} currents have previously been described²¹. In brief, a 2 – 5 mM Ba^{2+} NaCl external recording solution was used (in mM: 110 NaCl, 20 CsCl, 2–5 BaCl_2 , 1 MgCl_2 , 10 HEPES, 10 glucose, 0.2 TTX, pH 7.35–7.4, 300 mOsm l^{-1}).

The internal solution contained the following (in mM): 180 *N*-methyl-D-glucamine, 40 HEPES, 4 MgCl₂, 12 phosphocreatine, 2 Na₂ATP, 0.5 Na₂GTP, and 0.1 leupeptin, pH 7.2–3, with H₂SO₄, 265–270 mOsm l⁻¹. Calcium chelator was often omitted from the internal solution.

Tissue and single cell RT-PCR (scRT-PCR). In some experiments, dopaminergic neurons in the slice were aspirated, after electrophysiological characterization, and subjected to mRNA profiling using procedures similar to those previously described⁴⁹. Acutely isolated neurons were aspirated into sterilized glass pipettes containing nominally RNase-free patch solution or diethylpyrocarbonate-treated water and 0.8 U µl⁻¹ SUPERase-In (Ambion). Sterile gloves were worn during the procedure to minimize RNase contamination. After aspiration, the contents of the pipette were ejected into 0.6 ml presiliconized tubes (Midwest Scientific) containing a reverse transcription mix. This mix contained 0.7 µl of Superase-In (20 U µl⁻¹), 1.9 µl of diethylpyrocarbonate-treated water, 1 µl of dNTPs (10 mM), 0.7 µl of BSA (143 ng µl⁻¹), and 0.7 µl of oligo-dT (0.5 µg µl⁻¹). The mixture was heated to 65 °C for 5 min to linearize mRNA and then placed on ice for at least 1 min. Single-strand cDNA was synthesized from the cellular mRNA by adding 2 µl of 10× PCR buffer, 4 µl of MgCl₂ (25 mM), 2 µl of DTT (0.1 M), 1 µl of RNase out (40 U µl⁻¹), and 6 µl of diethylpyrocarbonate-treated water. This mixture was then incubated at 42 °C for 2 min. After the initial incubation, 0.7 µl of Superscript II (50 U µl⁻¹) was added, and the mixture was kept at 42 °C for an additional 50 min. The reaction was terminated by heating to 70 °C for 15 min. The RNA strand in the RNA–DNA hybrid was then removed by adding 0.5 µl of RNase H (2 U µl⁻¹) and incubating at 37 °C for 20 min. All reagents except Superase-In (Ambion) were obtained from Invitrogen. Single-cell cDNA was amplified using a conventional PCR approach with a programmable thermal cycler (MJ Research). PCR primers were developed from GenBank sequences with commercially available OLIGO 6.7.1 software (National Biosciences). Primers and reaction protocols for TH, calbindin D28k (CBN), Ca_v1.2 and Ca_v1.3 channels have been described previously^{21,58–60}. AC2 mRNA (GenBank accession NM153534) was detected with a pair of primers CCAGGAACCGAGCGTCAATA (position 2895) and ACTCCGGTGTCTGCTCCATCTACT (position 3100), which gave a PCR product of 228 bp. AC5 mRNA (GenBank accession NM001012765) was detected with a pair of primers GGCCTCAACGACTCCACCTATGA (position 3601) and CCTTGCCCTTGACCTTGACC (position 3921), which gave a PCR product of 340 bp. AC9 mRNA (GenBank accession NM009624) was detected with a pair of primers CGTATCCAGGTGAGCGAAGA (position 3655) and GAAGCCTTGTCGCAATACTGA (position 3897), which gave a PCR product of 263 bp. After amplification, PCR amplicons were labelled by ethidium bromide and separated by electrophoresis on agarose gels. Amplicons were of the expected size and sequence. RT-PCR was performed using procedures designed to minimize the chance of cross-contamination. Negative controls for contamination from extraneous DNA were run for every batch of neurons. Contamination from extraneous sources was checked for by eliminating the cellular template. The controls were consistently negative in these experiments.

Rotenone exposure and tissue preparation. Coronal midbrain slices (275 µm) were prepared as those for electrophysiological experiments. For rotenone exposure experiments, slices were incubated in ACSF for 2 h at 34 °C without or with isradipine (20 µM) followed by incubation with rotenone (300 nM) for 1 h at 34 °C. For control experiments, a sister slice from each animal was incubated in ACSF only for 3 h with and without isradipine (20 µM). Following incubation, slices were fixed in 4% paraformaldehyde in 0.1M phosphate buffer, pH 7.3–7.4 for 30 min at room temperature and cryoprotected in 30% sucrose. Slices were flat embedded in Tissue-Tek O.C.T. compound (4583, Sakura Finetek USA) and resectioned (at 30 µm) in the same plane as the parent slice on a cryostat (CM3000, Leica Instrument).

Immunocytochemistry in rotenone-treated slices. For light microscopy, sections were incubated with mouse anti-TH (1:3000; ImmunoStar) diluted in phosphate-buffered saline (PBS) containing 10% normal goat serum (NGS) and 0.05% Triton X-100 (Tx) for 24 h at 4 °C. After washes in PBS, the sections were incubated with biotinylated donkey anti-mouse IgG (1:500; Amersham) diluted in PBS-Tx containing 1% NGS at room temperature for 2 h and then reacted with avidin-biotin peroxidase complex (ABC-Elite kit, Vector Laboratories) at room temperature for 2 h. Bound peroxidase enzyme activity was revealed using Tris-buffered saline (pH 7.3) containing 0.025% 3,3'-diaminobenzidine tetrahydrochloride (DAB) and 0.003% hydrogen peroxide. Images were captured using a Spot RT CCD video camera (Diagnostics Instruments) mounted on the Nikon Eclipse 800 microscope (Nikon Instech Co.). For immunofluorescence double-labelling experiments, sections were incubated with rabbit anti-TH (1:300; Affinity BioReagents) and mouse anti-MAP2 (1:250; Sigma). After rinsing in PBS, the mixture of Alexa(488 and 568)-conjugated secondary antibodies (1:500; goat IgG, Invitrogen) were applied for 30 min. Sections were coverslipped with Vectashield (Vector Laboratories),

images were acquired at a Zeiss laser-scanning confocal microscope (LSM 510, Carl Zeiss International). As controls for immunostaining specificity, sections were incubated with the omission of either the primary or the secondary. Under these conditions, no specific immunostaining for the respective molecules was observed (data not shown).

Dendritic processes quantification in rotenone treated slices. To quantify density of TH-positive processes within the substantia nigra (–2.92 mm to –3.08 mm relative to bregma), three sections were sampled from three mice for each experimental and control condition. Substantia nigra pars reticulata was subdivided into medial, intermediate and lateral parts. For each section, the microscope field containing intermediate part was captured using a 40× objective (oil immersion, NA 1.3, Nikon Instech Co.) objective. A grid with subdivisions of ~36 µm × 36 µm was superimposed onto captured images in Adobe Photoshop CS. The grid was aligned parallel to the TH-positive cell bodies in SNc. The number of TH-positive processes that traversed a whole square was counted using every other square. Processes that were interrupted were not counted. The density of counted dendrites was calculated and statistically analysed.

Stereological analysis of TH staining in MPTP treated mice. Five days after chronic MPTP treatment, mice were anesthetized (135 mg kg⁻¹ sodium pentobarbital) and perfused transcardially with 3% paraformaldehyde in 0.1 M phosphate buffer, pH 7.3. Each brain was removed, sunk overnight in a sucrose buffer, and then cut coronally on a freezing microtome at 50 µm. Sections were incubated in monoclonal mouse anti-TH antibodies (1:2,000) using the reagents from the Mouse-on-Mouse kit (Vector Laboratories). Incubation in biotinylated anti-mouse IgG and further reacted with reagents from the ABC Elite kit (Vector Laboratories). Sections were counterstained with cresyl violet. Neurons in the SNc were counted stereologically for 4–7 mice as described elsewhere⁶¹. Briefly, analysis was carried out with a Nikon E400 microscope equipped with a 100× oil objective and high numerical aperture (1.35), and motorized stage in three axes (x, y, z). After a random start, every fourth, 50-µm-thick section was analysed following optical dissector rules using dedicated software (Stereo Investigator, MicroBrightField). The total number of Nissl-positive/TH-immunoreactive neurons in each dissector volume was estimated. The estimated total number of neurons in the SNc for each animal was obtained by doubling the number counted on one side of the brain.

Behavioural analysis of MPTP treated mice. Two days after chronic MPTP treatment, mice were subjected to a grid test to assess skilled forepaw use and coordination that are dependent on dopaminergic neurons⁶². Mice were placed in the centre of a wire mesh grid, which was then inverted, suspending the mouse 20 cm above the floor. Three trials of 60 s were videotaped in which mice were allowed to move freely while upside down; records were evaluated for forepaw stepping faults and total steps by an investigator blind to treatment condition. A ratio of forepaw foot faults/total steps was then calculated and data pooled for each group.

Analytic methods for MPTP metabolism. Isradipine- and placebo-pellet implantations were done similarly to that used for chronic MPTP regimen. Animals from both groups were injected with MPTP (30 mg kg⁻¹, i.p.). 90 min later, animals were killed by cervical dislocation and decapitation. Brains were removed and hemi-sectioned. Striata from each hemisphere were microdissected, placed into preweighed tubes and snap-frozen on dry ice. Samples were then stored at –80 °C before subsequent extraction processes. Individual striatum was homogenized in 1.5 ml microtubes with disposable polypropylene pestles (Kimble-Kontes). An aliquot (100 µl) of homogenizing buffer containing perchloric acid (0.4 M), sodium metabisulphite (0.1%, w/v), EDTA (0.01%, w/v), cysteine (0.1%, w/v) was used per 10 mg of tissue. After mechanical homogenization, samples were sonicated for 10 min at 4 °C, resuspended with buffer at 1 to 10 dilution. 200 µl of the resuspended tissue homogenates were centrifuged in 1.5 ml microtubes for 10 min (10,000 r.p.m.; 4 °C). Supernatant was retained for subsequent analysis of striatal MPP+ content using a liquid chromatography–tandem mass spectrometry system (LC/MS/MS) composed of a triple quadrupole mass spectrometer (API 3000, Applied Biosystems) equipped with a HPLC system (1100 series, Agilent Technologies). Standard curves were run each time along with samples as internal control and remain linear throughout the entire concentration range (1–1,000 ng ml⁻¹). Two independent replicates of each brain sample were analysed and averaged, quantification was only deemed valid if the error is less than 5%.

TH+ cell count from unilateral 6-OHDA lesioned animals. To quantify 6-OHDA toxicity on SNc neurons from placebo- and isradipine-treated animals, sections containing SNc from different rostrocaudal levels were immunohistochemically processed for TH. Images were captured using a 40× objective (oil immersion, NA 1.3, Nikon Instech Co.). The number of TH+ neurons in mid-lateral SNc from both control and lesioned side was counted and statistically analysed.

Data analysis and statistics methods. Curve fitting (using a least-squares criterion) and data analysis were done with IgorPro 5.05 (WaveMetrics). Numerical data are expressed and sample statistics are given as mean \pm s.e.m. (for samples ≥ 10) or median (interquartile range), for smaller samples. Distribution free box plots were used for graphic presentation of the data in most cases and exclusively when the sample size was less than 10. In data presented as box plots, the central line represents the median, the edges of the box represent the interquartile range, and the 'whisker lines' show the extent of the overall distribution, excluding outliers (points $> \pm 1.5 \times$ intraquartile range), which are shown as circles. Small, non-matched samples were analysed with non-parametric Kruskal–Wallis ANOVA, using a probability (P) threshold of 0.05. For stereological and behavioural analysis, data groups were compared with a one-way ANOVA followed by pairwise multiple comparison procedures (Student–Newman–Keuls method).

Computer simulation. SNc dopaminergic neurons were modelled with NEURON, version 5.9 (ref. 63). The model incorporated biophysically accurate mod files describing Na^+ , Ca^{2+} , K^+ and cation (HCN) channels known to contribute to pacemaking. Axial resistivity was assumed to be $70 \Omega \text{cm}^{-1}$. Intracellular Ca^{2+} diffusion, buffering and extrusion were based on mechanisms bundled with NEURON. SK channel gating was modelled as previously described⁶⁴. HCN modelling was based upon ref. 65; KCNQ modelling was as previously described⁶⁶. The model used a single compartment (soma: length, 15 μm ; diameter, 15 μm). Channels in the somatic membrane (densities in S cm^{-2}) were: Na^+ (10^{-2}), Kv2 (1.5×10^{-4}), KCNQ (7×10^{-7}), SK (5×10^{-6}), HCN (7×10^{-5}), $\text{Ca}_v1.3$ (5×10^{-5}), $\text{Ca}_v2.1$ (2.5×10^{-5}), leak ($-60 E_{\text{rev}}$, 5×10^{-6}). Mod files for Na , Kv2 , Kv7 , KCNQ , SK , HCN , $\text{Ca}_v1.3$, CaP and Ca handling are available on request; the model will be posted on the NEURON DB website (<http://senselab.med.yale.edu/senselab/modeldb>).

Reagents and channel ligands. All reagents were obtained from Sigma-Aldrich, except KMeSO_4 (MP Biomedicals), $\text{Na}_2\text{-GTP}$ (Roche Diagnostics), isradipine, ZD 7288 (Tocris Cookson), TTX (Alomone Labs), rolipram, *dd*-adenosine, NKY80 (Calbiochem). Drugs were dissolved as stock solutions in water, DMSO or anhydrous methanol (99.9%) and aliquoted and frozen at -30°C before use. Each of the drugs was diluted in ACSF immediately before the experiment. When used, the final concentration of DMSO or methanol in the external solution was always $<0.01\%$.

51. Platzter, J. *et al.* Congenital deafness and sinoatrial node dysfunction in mice lacking class D L-type Ca^{2+} channels. *Cell* **102**, 89–97 (2000).
52. Nelson, E. L., Liang, C. L., Sinton, C. M. & German, D. C. Midbrain dopaminergic neurons in the mouse: Computer-assisted mapping. *J. Comp. Neurol.* **369**, 361–371 (1996).
53. Paxinos, G. & Franklin, K. B. J. *The Mouse Brain in Stereotaxic Coordinates* (Elsevier Academic, San Diego, 2004).
54. Grace, A. A. & Bunney, B. S. The control of firing pattern in nigral dopamine neurons: Single spike firing. *J. Neurosci.* **4**, 2866–2876 (1984).
55. Grace, A. A. & Onn, S. P. Morphology and electrophysiological properties of immunocytochemically identified rat dopamine neurons recorded *in vitro*. *J. Neurosci.* **9**, 3463–3481 (1989).
56. Yung, W. H., Hausser, M. A. & Jack, J. J. Electrophysiology of dopaminergic and non-dopaminergic neurones of the guinea-pig substantia nigra pars compacta *in vitro*. *J. Physiol. (Lond.)* **436**, 643–667 (1991).
57. Maurice, N. *et al.* D2 dopamine receptor-mediated modulation of voltage-dependent Na^+ channels reduces autonomous activity in striatal cholinergic interneurons. *J. Neurosci.* **24**, 10289–10301 (2004).
58. Tkatch, T., Baranaskas, G. & Surmeier, D. J. Basal forebrain neurons adjacent to the globus pallidus co-express GABAergic and cholinergic marker mRNAs. *Neuroreport* **9**, 1935–1939 (1998).
59. Song, W. J. *et al.* Somatodendritic depolarization-activated potassium currents in rat neostriatal cholinergic interneurons are predominantly of the A type and attributable to coexpression of Kv4.2 and Kv4.1 subunits. *J. Neurosci.* **18**, 3124–3137 (1998).
60. Surmeier, D. J., Song, W. J. & Yan, Z. Coordinated expression of dopamine receptors in neostriatal medium spiny neurons. *J. Neurosci.* **16**, 6579–6591 (1996).
61. Mandir, A. S. *et al.* Poly(ADP-ribose) polymerase activation mediates 1-methyl-4-phenyl-1, 2,3,6-tetrahydropyridine (MPTP)-induced parkinsonism. *Proc. Natl Acad. Sci. USA* **96**, 5774–5779 (1999).
62. Meredith, G. E. & Kang, U. J. Behavioral models of Parkinson's disease in rodents: A new look at an old problem. *Mov. Disord.* **21**, 1595–1606 (2006).
63. Hines, M. L. & Carnevale, N. T. NEURON: A tool for neuroscientists. *Neuroscientist* **7**, 123–135 (2001).
64. Migliore, M., Cook, E. P., Jaffe, D. B., Turner, D. A. & Johnston, D. Computer simulations of morphologically reconstructed CA3 hippocampal neurons. *J. Neurophysiol.* **73**, 1157–1168 (1995).
65. Wang, J., Chen, S., Nolan, M. F. & Siegelbaum, S. A. Activity-dependent regulation of HCN pacemaker channels by cyclic AMP: Signaling through dynamic allosteric coupling. *Neuron* **36**, 451–461 (2002).

66. Shen, W., Hamilton, S. E., Nathanson, N. M. & Surmeier, D. J. Cholinergic suppression of KCNQ channel currents enhances excitability of striatal medium spiny neurons. *J. Neurosci.* **25**, 7449–7458 (2005).

Genome-wide association study identifies novel breast cancer susceptibility loci

Douglas F. Easton¹, Karen A. Pooley², Alison M. Dunning², Paul D. P. Pharoah², Deborah Thompson¹, Dennis G. Ballinger³, Jeffery P. Struwing⁴, Jonathan Morrison², Helen Field², Robert Luben⁵, Nicholas Wareham⁵, Shahana Ahmed², Catherine S. Healey², Richard Bowman⁶, the SEARCH collaborators^{2*}, Kerstin B. Meyer⁷, Christopher A. Haiman⁸, Laurence K. Kolonel⁹, Brian E. Henderson⁸, Loic Le Marchand⁹, Paul Brennan¹⁰, Suleeporn Sangrajrang¹¹, Valerie Gaborieau¹⁰, Fabrice Odefrey¹⁰, Chen-Yang Shen¹², Pei-Ei Wu¹², Hui-Chun Wang¹², Diana Eccles¹³, D. Gareth Evans¹⁴, Julian Peto¹⁵, Olivia Fletcher¹⁶, Nichola Johnson¹⁶, Sheila Seal¹⁷, Michael R. Stratton^{17,18}, Nazneen Rahman¹⁷, Georgia Chenevix-Trench¹⁹, Stig E. Bojesen²⁰, Børge G. Nordestgaard²⁰, Christen K. Axelsson²¹, Montserrat Garcia-Closas²², Louise Brinton²², Stephen Chanock²³, Jolanta Lissowska²⁴, Beata Peplonska²⁵, Heli Nevanlinna²⁶, Rainer Fagerholm²⁶, Hannaleena Eerola^{26,27}, Daehee Kang²⁸, Keun-Young Yoo^{28,29}, Dong-Young Noh²⁸, Sei-Hyun Ahn³⁰, David J. Hunter^{31,32}, Susan E. Hankinson³², David G. Cox³¹, Per Hall³³, Sara Wedren³³, Jianjun Liu³⁴, Yen-Ling Low³⁴, Natalia Bogdanova^{35,36}, Peter Schürmann³⁶, Thilo Dörk³⁶, Rob A. E. M. Tollenaar³⁷, Catharina E. Jacobi³⁸, Peter Devilee³⁹, Jan G. M. Klijn⁴⁰, Alice J. Sigurdson⁴¹, Michele M. Doody⁴¹, Bruce H. Alexander⁴², Jinghui Zhang⁴, Angela Cox⁴³, Ian W. Brock⁴³, Gordon MacPherson⁴³, Malcolm W. R. Reed⁴⁴, Fergus J. Couch⁴⁵, Ellen L. Goode⁴⁵, Janet E. Olson⁴⁵, Hanne Meijers-Heijboer^{46,47}, Ans van den Ouweland⁴⁷, André Uitterlinden⁴⁸, Fernando Rivadeneira⁴⁸, Roger L. Milne⁴⁹, Gloria Ribas⁴⁹, Anna Gonzalez-Neira⁴⁹, Javier Benitez⁴⁹, John L. Hopper⁵⁰, Margaret McCredie⁵¹, Melissa Southey⁵⁰, Graham G. Giles⁵², Chris Schroen⁵³, Christina Justenhoven⁵⁴, Hiltrud Brauch⁵⁴, Ute Hamann⁵⁵, Yon-Dschun Ko⁵⁶, Amanda B. Spurdle¹⁹, Jonathan Beesley¹⁹, Xiaoqing Chen¹⁹, kConFab^{57*}, AOCs Management Group^{19,57*}, Arto Mannermaa^{58,59}, Veli-Matti Kosma^{58,59}, Vesa Kataja^{58,60}, Jaana Hartikainen^{58,59}, Nicholas E. Day⁵, David R. Cox³ & Bruce A. J. Ponder^{2,7}

Breast cancer exhibits familial aggregation, consistent with variation in genetic susceptibility to the disease. Known susceptibility genes account for less than 25% of the familial risk of breast cancer, and the residual genetic variance is likely to be due to variants conferring more moderate risks. To identify further susceptibility alleles, we conducted a two-stage genome-wide association study in 4,398 breast cancer cases and 4,316 controls, followed by a third stage in which 30 single nucleotide polymorphisms (SNPs) were tested for confirmation in 21,860 cases and 22,578 controls from 22 studies. We used 227,876 SNPs that were estimated to correlate with 77% of known common SNPs in Europeans at $r^2 > 0.5$. SNPs in five novel independent loci exhibited strong and consistent evidence of association with breast cancer ($P < 10^{-7}$). Four of these contain plausible causative genes (*FGFR2*, *TNRC9*, *MAP3K1* and *LSP1*). At the second stage, 1,792 SNPs were significant at the $P < 0.05$ level compared with an estimated 1,343 that would be expected by chance, indicating that many additional common susceptibility alleles may be identifiable by this approach.

Breast cancer is about twice as common in the first-degree relatives of women with the disease as in the general population, consistent with variation in genetic susceptibility to the disease¹. In the 1990s, two major susceptibility genes for breast cancer, *BRCA1* and *BRCA2*, were identified^{2,3}. Inherited mutations in these genes lead to a high risk of breast and other cancers⁴. However, the majority of multiple case breast cancer families do not segregate mutations in these genes. Subsequent genetic linkage studies have failed to identify further major breast cancer genes⁵. These observations have led to the proposal that breast cancer susceptibility is largely 'polygenic': that is, susceptibility is conferred by a large number of loci, each with a small effect on breast cancer risk⁶. This model is consistent with the observed patterns of familial aggregation of breast cancer⁷. However,

progress in identifying the relevant loci has been slow. As linkage studies lack power to detect alleles with moderate effects on risk, large case-control association studies are required. Such studies have identified variants in the DNA repair genes *CHEK2*, *ATM*, *BRIP1* and *PALB2* that confer an approximately twofold risk of breast cancer, but these variants are rare in the population^{8–14}. A recent study has shown that a common coding variant in *CASP8* is associated with a moderate reduction in breast cancer risk¹⁵. After accounting for all the known breast cancer loci, more than 75% of the familial risk of the disease remains unexplained¹⁶.

Recent technological advances have provided platforms that allow hundreds of thousands of SNPs to be analysed in association studies, thus providing a basis for identifying moderate risk alleles without

Affiliations of the above authors are given at the end of the paper.

*Lists of consortia participants and affiliations appear after author affiliations.

prior knowledge of position or function. It has been estimated that there are 7 million common SNPs in the human genome (with minor allele frequency, m.a.f., $>5\%$)¹⁷. However, because recombination tends to occur at distinct 'hot-spots', neighbouring polymorphisms are often strongly correlated (in 'linkage disequilibrium', LD) with each other. The majority of common genetic variants can therefore be evaluated for association using a few hundred thousand SNPs as tags for all the other variants¹⁸. We aimed to identify further breast cancer susceptibility loci in a three-stage association study¹⁹. In the first stage, we used a panel of 266,722 SNPs, selected to tag known common variants across the entire genome¹⁸. These SNPs were genotyped in 408 breast cancer cases and 400 controls from the UK; data were analysed for 390 cases and 364 controls genotyped for $\geq 80\%$ of the SNPs. The cases were selected to have a strong family history of breast cancer, equivalent to at least two affected female first-degree relatives, because such cases are more likely to carry susceptibility alleles²⁰. Initially, we analysed 227,876 SNPs (85%) with genotypes on at least 80% of the subjects. We estimate that these SNPs are correlated with 58% of common SNPs in the HapMap CEPH/CEU (Utah residents with ancestry from northern and western Europe) samples at $r^2 > 0.8$, and 77% at $r^2 > 0.5$ (mean $r^2 = 0.75$; see Supplementary Fig. 1) (<http://www.hapmap.org/>)²¹. As expected, coverage was strongly related to m.a.f.: 70% of SNPs with m.a.f. $> 10\%$ were tagged at $r^2 > 0.8$, compared with 23% of SNPs with m.a.f. 5–10%. The main analyses were restricted to 205,586 SNPs that had a call rate of 90% and whose genotype distributions did not differ from Hardy–Weinberg equilibrium in controls (at $P < 10^{-5}$).

For the second stage we selected 12,711 SNPs, approximately 5% of those typed in stage 1, on the basis of the significance of the difference in genotype frequency between cases and controls. These SNPs were

then genotyped in a further 3,990 invasive breast cancer cases and 3,916 controls from the SEARCH study, using a custom-designed oligonucleotide array. In the main analyses, we considered 10,405 SNPs with call rate of $>95\%$ that did not deviate from Hardy–Weinberg equilibrium in controls.

Comparison of the observed and expected distribution of test statistics showed some evidence for an inflation of the test statistics in both stage 1 (inflation factor $\lambda = 1.03$, 95% confidence interval (CI) 1.02–1.04) and stage 2 ($\lambda = 1.06$, 95% CI 1.04–1.12), based on the 90% least significant SNPs (Fig. 1). Possible explanations for this inflation include population stratification, cryptic relatedness among subjects, and differential genotype calling between cases and controls. There was evidence for an excess of low call rate SNPs among the most significant SNPs ($P < 0.01$) in stage 1, but not in stage 2, suggesting that some of this effect is a genotyping artefact (Supplementary Table 1). However, the inflation was still present among SNPs with call rate $>99\%$ in both cases and controls, possibly reflecting population substructure. We computed 1 degree of freedom (d.f.) association tests for each SNP, combining stages 1 and 2. After adjustment for this inflation by the genomic control method²², we observed more associations than would have been expected by chance at $P < 0.05$ (Table 1). One SNP (dbSNP rs2981582) was significant at the $P < 10^{-7}$ level that has been proposed as appropriate for genome-wide studies²³.

In the third stage, to establish whether any SNPs were definitely associated with risk, we tested 30 of the most significant SNPs in 22 additional case-control studies, comprising 21,860 cases of invasive breast cancer, 988 cases of carcinoma *in situ* (CIS) and 22,578 controls (Supplementary Table 2). Six SNPs showed associations in stage 3 that were significant at $P \leq 10^{-5}$ with effects in the same direction as in stages 1 and 2 (Table 2, Supplementary Table 3, and Fig. 2). All these SNPs reached a combined significance level of $P < 10^{-7}$ (ranging from 2×10^{-76} to 3×10^{-9}). Of these six SNPs, five were within genes or LD blocks containing genes. SNP rs2981582 lies in intron 2 of *FGFR2* (also known as *CEK3*), which encodes the fibroblast growth factor receptor 2. SNPs rs12443621 and rs8051542 are both located in an LD block containing the 5' end of *TNRC9* (also known as *TOX3*), a gene of uncertain function containing a tri-nucleotide repeat motif, as well as the hypothetical gene, *LOC643714*. SNP rs889312 lies in an LD block of approximately 280 kb that contains *MAP3K1* (also known as *MEKK*), which encodes the signalling protein mitogen-activated protein kinase kinase 1, in addition to two other genes: *MGC33648* and *MIER3*. SNP rs3817198 lies in intron 10 of *LSP1* (also known as *WP43*), encoding lymphocyte-specific protein 1, an F-actin bundling cytoskeletal protein expressed in haematopoietic and endothelial cells. A further SNP, rs2107425, located just 110 kilobases (kb) from rs3817198, was also identified (overall $P = 0.00002$). rs2107425 is within the *H19* gene, an imprinted maternally expressed untranslated messenger RNA closely involved in regulation of the insulin growth factor gene, *IGF2*. In stage 3, however, rs2107425 was only weakly significant after adjustment for rs3817198 by logistic regression ($P = 0.06$). This suggests that the association with breast cancer risk may be driven by variants in *LSP1* rather than in *H19*. The sixth SNP reaching a combined $P < 10^{-7}$ was rs13281615, which lies on 8q. It is correlated with SNPs in a 110 kb LD block that contains no known

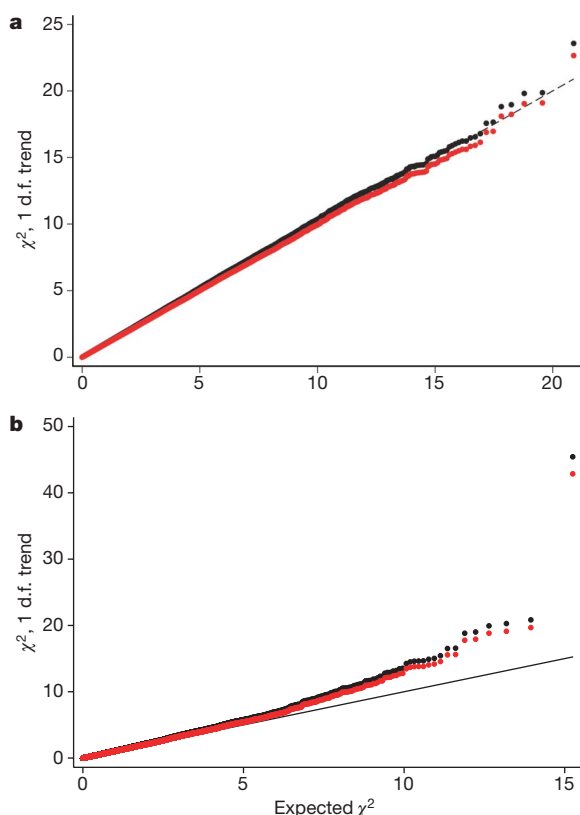


Figure 1 | Quantile-quantile plots for the test statistics (Cochran-Armitage 1 d.f. χ^2 trend tests) for stages 1 and 2. a, Stage 1; b, stage 2. Black dots are the uncorrected test statistics. Red dots are the statistics corrected by the genomic control method ($\lambda = 1.03$ for stage 1, $\lambda = 1.06$ for stage 2). Under the null hypothesis of no association at any locus, the points would be expected to follow the black line.

Table 1 | Number of significant associations after stage 2

Level of significance	Observed	Observed adjusted*	Expected	Ratio
0.01–0.05	1,239	1,162	934.3	1.24
0.001–0.01	574	517	347.6	1.49
0.0001–0.001	112	88	53.3	1.65
0.00001–0.0001	16	12	7.0	1.71
<0.00001	15	13	0.96	13.5
All $P < 0.05$	1,956	1,792	1,343.2	1.33

Observed numbers of SNPs associated with breast cancer after stage 2, by level of significance, before and after adjustment for population stratification, and expected numbers under the null hypothesis of no association.

* Adjusted for inflation of the test statistic by the genomic control method.

Table 2 | Summary of results for eleven SNPs selected for stage 3 that showed evidence of an association with breast cancer

rs Number	Gene	Position*	m.a.f.†	Per allele OR (95% CI)	HetOR (95% CI)	HomOR (95% CI)	P-trend		
							Stages 1 and 2	Stage3	Combined
rs2981582	<i>FGFR2</i>	10q 123342307	0.38 (0.30)	1.26 (1.23–1.30)	1.23 (1.18–1.28)	1.63 (1.53–1.72)	4×10^{-16}	5×10^{-62}	2×10^{-76}
rs12443621	<i>TNRC9/</i> <i>LOC643714</i>	16q 51105538	0.46 (0.60)	1.11 (1.08–1.14)	1.14 (1.09–1.20)	1.23 (1.17–1.30)	10^{-7}	9×10^{-14}	2×10^{-19}
rs8051542	<i>TNRC9/</i> <i>LOC643714</i>	16q 51091668	0.44 (0.20)	1.09 (1.06–1.13)	1.10 (1.05–1.16)	1.19 (1.12–1.27)	4×10^{-6}	4×10^{-8}	10^{-12}
rs889312	<i>MAP3K1</i>	5q 56067641	0.28 (0.54)	1.13 (1.10–1.16)	1.13 (1.09–1.18)	1.27 (1.19–1.36)	4×10^{-6}	3×10^{-15}	7×10^{-20}
rs3817198	<i>LSP1</i>	11p 1865582	0.30 (0.14)	1.07 (1.04–1.11)	1.06 (1.02–1.11)	1.17 (1.08–1.25)	8×10^{-6}	10^{-5}	3×10^{-9}
rs2107425	<i>H19</i>	11p 1977651	0.31 (0.44)	0.96 (0.93–0.99)	0.94 (0.90–0.98)	0.95 (0.89–1.01)	7×10^{-6}	0.01	2×10^{-5}
rs13281615		8q 128424800	0.40 (0.56)	1.08 (1.05–1.11)	1.06 (1.01–1.11)	1.18 (1.10–1.25)	2×10^{-7}	6×10^{-7}	5×10^{-12}
rs981782		5p 45321475	0.47 (0.37)	0.96 (0.93–0.99)	0.96 (0.92–1.01)	0.92 (0.87–0.97)	8×10^{-5}	0.003	9×10^{-6}
rs30099		5q 52454339	0.08 (0.39)	1.05 (1.01–1.10)	1.06 (1.00–1.11)	1.09 (0.96–1.24)	0.003	0.02	0.001
rs4666451		2p 19150424	0.41 (0.04)	0.97 (0.94–1.00)	0.98 (0.93–1.02)	0.93 (0.87–0.99)	5×10^{-6}	0.04	6×10^{-5}
rs3803662‡	<i>TNRC9/</i> <i>LOC643714</i>	16q 51143842	0.25 (0.60)	1.20 (1.16–1.24)	1.23 (1.18–1.29)	1.39 (1.26–1.45)	3×10^{-12}	10^{-26}	10^{-36}

OR, odds ratio; HetOR, odds ratio in heterozygotes; HomOR, odds ratio in rare homozygotes (relative to common homozygotes); CI, confidence interval.

* Build 36.2 position.

† Minor allele frequency in SEARCH (UK) study. Combined allele frequency from three Asian studies in *italics*.

‡ rs3803662 was not part of the initial tag SNP set but identified as a result of fine-scale mapping of the *TNRC9/LOC643714* locus and typed in the stage 2 and stage 3 sets (but not the stage 1 set).

genes. The basis of this association therefore remains obscure. This SNP is approximately 130 kb proximal to rs1447295, 60 kb proximal to rs6983267 and 230 kb distal to rs16901979, recently shown to be associated with prostate cancer^{24–26}.

In addition to the seven SNPs described above, there was evidence of association among the remaining 23 SNPs (global $P = 0.001$ in stage 3). In particular, three SNPs showed some evidence of association in stage 3 ($P < 0.05$, in each case in the same direction as in stages 1 and 2; Table 2). SNPs rs981782 and rs30099 both lie in the centromeric region of chromosome 5. rs4666451 lies on 2p, a region for which some evidence of linkage to breast cancer in families has been reported⁵. The 20 other SNPs showed no evidence of association in stage 3 (global $P = 0.11$), suggesting that most of these associations from stages 1 and 2 were false positives.

FGFR2

The most significantly associated SNP, rs2981582, lies within a 25 kb LD block almost entirely within intron 2 of *FGFR2*. We found no evidence of association with SNPs elsewhere in the gene (Fig. 3a). In an attempt to identify a causal variant, we first identified the 19 common variants ($m.a.f. > 0.05$) in this block from HapMap CEU data. These were tagged ($r^2 > 0.8$) by 7 SNPs including rs2981582. The additional tag SNPs were genotyped in the SEARCH study cases and controls. Multiple logistic regression analysis of these variants found no additional evidence for association after adjusting for rs2981582. Haplotype analysis of these 7 SNPs indicated that multiple haplotypes carrying the minor (*a*) allele of rs2981582 were associated with an increased risk of breast cancer, implying that the association was being driven by rs2981582 itself or a variant strongly correlated with it (Supplementary Table 4).

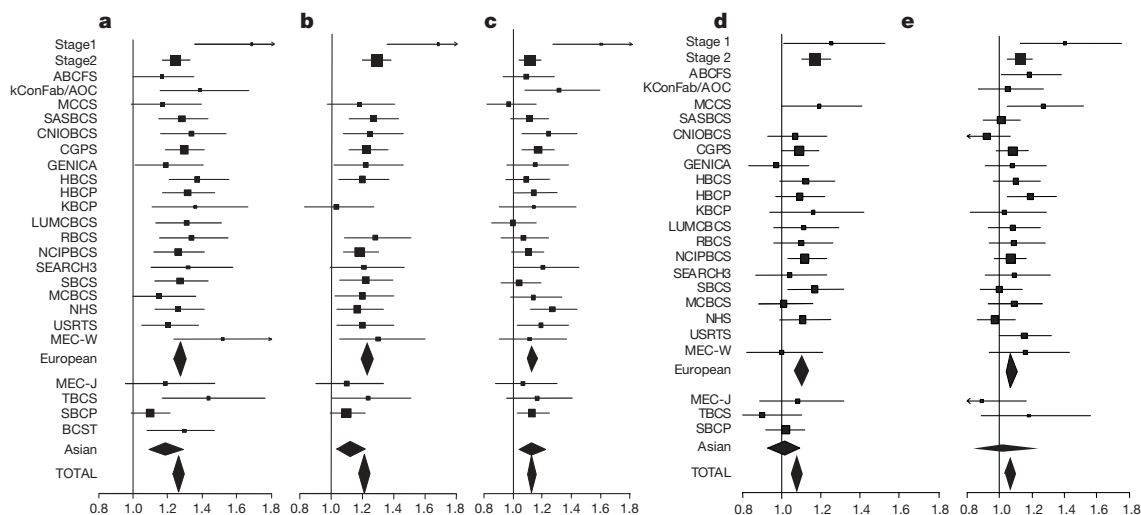


Figure 2 | Forest plots of the per-allele odds ratios for each of the five SNPs reaching genome-wide significance. a, rs2981582; b, rs3803662; c, rs889312; d, rs13281615; and e, rs3817198. The x-axis gives the per-allele odds ratio. Each row represents one study (see Supplementary Table 2), with summary odds ratios for all European and all Asian studies, and all studies combined.

The area of the square for each study is proportional to the inverse of the variance of the estimate. Horizontal lines represent 95% confidence intervals. Diamonds represent the summary odds ratios, with 95% confidence intervals, based on the stage 3 studies only.

Resequencing of this region in 45 subjects of European origin identified 29 variants that were strongly correlated with rs2981582 ($r^2 > 0.6$) (<http://cgwb.nci.nih.gov>; Fig. 3b and Supplementary Tables 5–8). A subset of 14 variants tagged 27 of these in European ($r^2 > 0.95$) and Asian (Korean) samples ($r^2 > 0.86$). Two variants could not be genotyped reliably. This new tagging set was then genotyped in SEARCH and 3 studies from Asian populations; the Asian studies were included because the LD is weaker, providing greater power to resolve the causal variant (Fig. 3b, left panel). The strongest association was found with rs7895676. On the assumption that there is a single disease-causing allele, we calculated a likelihood for each variant. 21 SNPs (including rs2981582) had a likelihood ratio of $< 1/100$ relative to rs7895676, indicating that none of these are likely to be the causal variant (Supplementary Table 8). Six variants were too strongly correlated for their individual effects to be separated using a genetic epidemiological approach. Functional assays will be required to determine which is causally related to breast cancer risk.

Intron 2 of *FGFR2* shows a high degree of conservation in mammals, and contains several putative transcription-factor binding sites (<http://genomequebec.mcgill.ca/PReMod>)²⁷, some of which lie in close proximity to the relevant SNPs. We therefore speculate that the association with breast cancer risk is mediated through regulation of *FGFR2* expression. Of possible relevance is that only three of these variants (rs10736303, rs2981578 and rs35054928) are within sequences conserved across all placental mammals (Fig. 3c and

Supplementary Table 8). Of these, the disease associated allele of rs10736303 generates a putative oestrogen receptor (ER) binding site. rs35054928 lies immediately adjacent to a perfect POU domain protein octamer (Oct) binding site. However, multiple splice variants have been reported in *FGFR2*, and differential splicing might provide an alternative mechanism for the association. *FGFR2* is a receptor tyrosine kinase that is amplified and overexpressed in 5–10% of breast tumours^{28–30}. Somatic missense mutations of *FGFR2* that are likely to be implicated in cancer development have also been demonstrated in primary tumours and cell lines of multiple tumour types (<http://www.sanger.ac.uk/genetics/CGP/cosmic/>)^{30,31}.

TNRC9/LOC643714 locus

As two SNPs in the *TNRC9/LOC643714* locus, rs12443621 and rs8051542, both showed convincing evidence of association, we further evaluated this region by genotyping, in the SEARCH set, an additional 19 SNPs tagging 101 common variants within the entire *TNRC9* and *LOC643714* genes, based on the HapMap CEU data. SNPs tagging the coding region of *TNRC9* showed no evidence of association. The strongest association was observed with rs3803662, a synonymous coding SNP of *LOC643714* that lies 8 kb upstream of *TNRC9*. This SNP was therefore genotyped in the stage 3 set (Table 2). Logistic regression analysis indicated that rs3803662 exhibited a stronger association with disease than other SNPs, and the associations with other SNPs were non-significant after adjustment for rs3803662. These results suggest

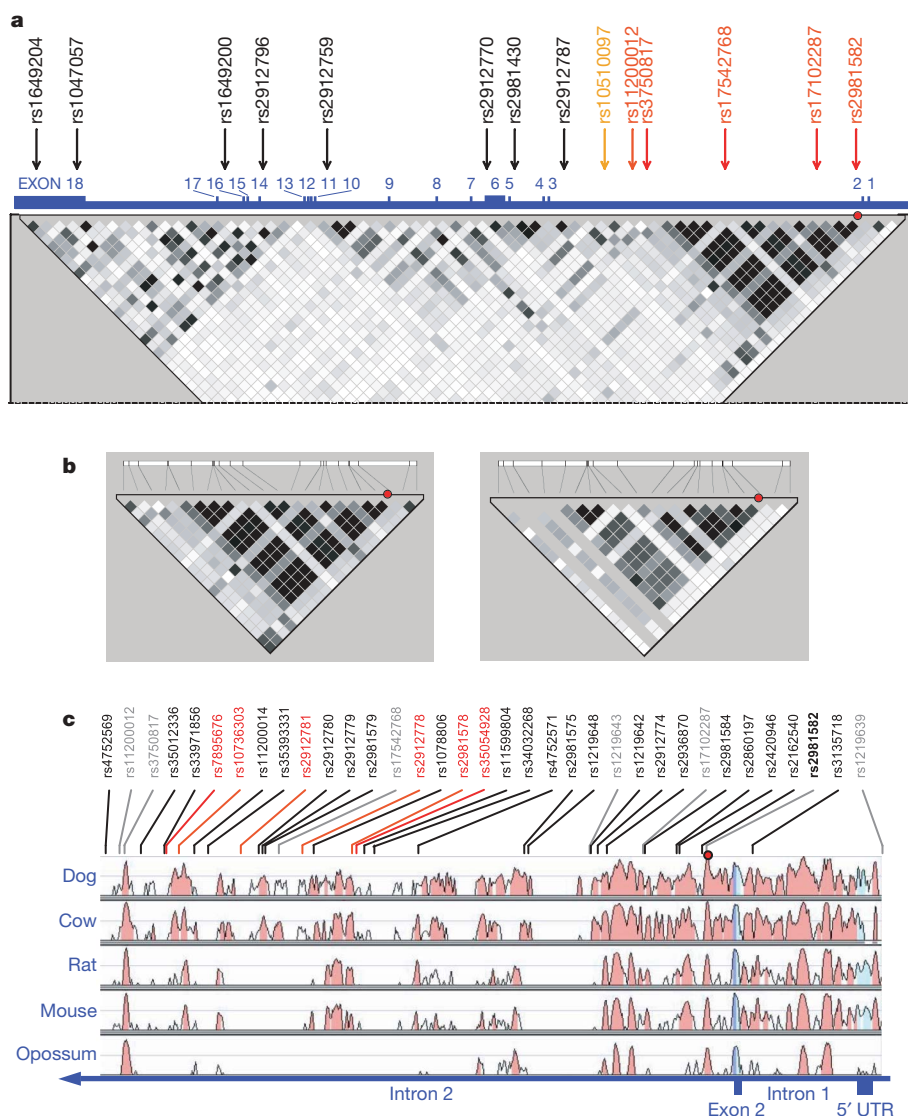


Figure 3 | The *FGFR2* locus. **a**, Map of the whole *FGFR2* gene, viewed relative to common SNPs on HapMap. The gene is 126 kb long and in reverse 3'–5' orientation on chromosome 10. Exon positions are illustrated with respect to the 67 SNPs with m.a.f. > 5% in HapMap CEU (therefore the map is not to physical scale). Numbered SNPs are those tested in the genome-wide study. SNPs in black were not significant in stage 1. Those in red were significant at $P < 0.0001$ after stage 2. rs10510097 (orange) was significant in stage 1, but failed quality control in stage 2 owing to deviation from Hardy–Weinberg equilibrium. Squares indicate pairwise r^2 on a greyscale (black = 1, white = 0). Red circle indicates rs2981582. **b**, Resequenced 32 kb region, shown relative to SNPs in CEU with m.a.f. > 5%, showing pairwise LD for SNPs in HapMap CEU (left panel) and JPT/CHB (right panel). Red circle indicates rs2981582, shown in bold black. **c**, Sequence conservation of 32 kb region in five species, relative to human sequence (<http://pipeline.lbl.gov/methods.shtml>)³⁵. Red circle indicates rs2981582. SNPs in grey are those used in the initial tagging of known common HapMap SNPs within the block. SNPs in black are correlated with rs2981582 with $r^2 > 0.6$ in European samples. Six SNPs in red were those consistent with being the causative variant on the basis of the genetic data (not excluded at odds of 100:1 relative to the SNP with the strongest association, rs7895676).

that the causal variant is closely correlated with rs3803662. Four SNPs in the HapMap CEU data (rs17271951, rs1362548, rs3095604 and rs4784227) that span *LOC643714* and the 5' regulatory regions of *TNRC9* are strongly correlated with rs3803662, and it therefore remains unclear in which gene the causative variant lies. *TNRC9* contains a putative HMG (high mobility group) box motif, suggesting that it might act as a transcription factor.

Pattern of risks

We assessed in more detail, in the stage 3 data, the pattern of the risks associated with the five independent SNPs that reached an overall $P < 10^{-7}$: rs2981582 (*FGFR2*), rs3803662 (*TNRC9/LOC643714*), rs889312 (*MAP3K1*), rs13281615 (8q) and rs3817198 (*LSP1*). For each of these five SNPs, the minor allele in Europeans was associated with an increased risk of breast cancer in a dose-dependent manner, with a higher risk of breast cancer in homozygous than in heterozygous carriers. Simple dominant and recessive models could be rejected for each SNP (all $P = 0.02$ or less). There was a marked difference in allele frequencies between populations, with the risk-associated alleles of rs8051542, rs889312 and rs13281615 being the major allele in Asian populations. The per allele odds ratio associated with rs2981582 was significantly smaller, though still elevated, in the Asian versus European populations ($P = 0.04$ for difference in odds ratio). This difference is consistent with the hypothesis that rs2981582 is not the functional variant at the *FGFR2* locus, and was not seen for SNPs exhibiting stronger evidence in the fine-scale mapping. No other evidence for heterogeneity in the per-allele odds ratio among studies was observed (Fig. 2).

Three of the SNPs (rs2981582, rs3803662 and rs889312) also showed evidence of association with breast CIS (Supplementary Table 9). For rs2981582 and rs3803662, the estimated odds ratios were greater for a diagnosis of breast cancer before age 40 years, but the trends by age were not statistically significant (Supplementary Table 10). There was evidence of an association with family history of breast cancer for three SNPs: for rs2981582 ($P = 0.02$), rs3803662 ($P = 0.03$) and rs13281615 ($P = 0.05$), the susceptibility allele was commoner in women with a first-degree relative with the disease than in those without (Supplementary Table 11). rs2981582 was also associated with bilaterality ($P = 0.02$). The associations with family history and bilaterality are to be expected for susceptibility loci, and are similar to previous observations for alleles in *CHEK2* and *ATM* (refs 10, 12, 14).

Discussion

This study has identified five novel breast cancer susceptibility loci, and demonstrated conclusively that some of the variation in breast cancer risk is due to common alleles. None of the loci we identified had been previously reported in association studies. Most previously identified breast cancer susceptibility genes are involved in DNA repair, and many association studies in breast cancer have concentrated on genes in DNA repair and sex hormone synthesis and metabolism pathways. None of the associations reported here appear to relate to genes in these pathways. It is notable that three of the five loci contain genes related to control of cell growth or to cell signalling, but only one (*FGFR2*) had a clear prior relevance to breast cancer. These results should, therefore, open up new avenues for basic research.

Our results emphasize the critical importance of study size in genetic association studies. It is notable that none of the confirmed associations reached genome-wide significance after stage 1 and only one reached this level after stage 2. As most common cancers have similar familial relative risks to breast cancer, it is likely that similarly large studies will be required to identify common alleles for other cancers. The fine-scale mapping of the *FGFR2* locus demonstrates that, even with a clear association, identification of the causative variant can be extremely problematic. However, the use of studies from multiple populations with different patterns of LD can substantially reduce the number of variants that need to be subjected to functional analysis.

As these susceptibility alleles are very common, a high proportion of the general population are carriers of at-risk genotypes. For example,

approximately 14% of the UK population and 19% of UK breast cancer cases are homozygous for the rare allele at rs2981582. On the other hand, the increased risks associated with these alleles are relatively small—on the basis of UK population rates, the estimated breast cancer risk by age 70 years for rare homozygotes at rs2981582 is 10.5%, compared to 6.7% in heterozygotes and 5.5% in common homozygotes. At this stage, it is unlikely that these SNPs will be appropriate for predictive genetic testing, either alone or in combination with each other. However, as further susceptibility alleles are identified, a combination of such alleles together with other breast cancer risk factors may become sufficiently predictive to be important clinically.

On the basis of the relative risk estimates from stage 3, and assuming that the five most significant loci interact multiplicatively on disease risk, these loci explain an estimated 3.6% of the excess familial risk of breast cancer. On the basis of our staged design and the estimated distribution of linkage disequilibrium between the typed SNPs and those in HapMap, we estimate that the power to identify the five most significant associations at $P < 10^{-7}$ (rs2981582, rs3803662, rs889312, rs13281615 and rs3817198) was 93%, 71%, 25%, 3% and 1% respectively. These estimates are uncertain, notably because the true coverage of HapMap SNPs is unknown. Nevertheless, these calculations indicate that the power to detect the two strongest associations was high, and suggest that there are likely to be few other common variants with a similar effect on variation in breast cancer risk to rs2981582. In contrast, the low power to detect rs13281615 and rs3817198 suggests that these variants may represent a much larger class of loci, each explaining of the order of 0.1% of the familial risk of breast cancer. An example of such a locus is provided by *CASP8* D302H, which showed strong evidence of association in a previous large study¹⁵. This SNP was tested in stage 1, but the association was missed because it did not reach the threshold for testing in stage 2. The excess of associations after stage 2 is also consistent with the existence of many such loci. In addition, because the coverage for SNPs with m.a.f. $< 10\%$ was low, many low frequency alleles may have been missed. The detection of further susceptibility loci will require genome-wide studies with more complete coverage and using larger numbers of cases and controls, together with the combination of results across multiple studies. The present study demonstrates that common susceptibility loci can be reliably identified, and that they may together explain an appreciable fraction of the genetic variance in breast cancer risk.

METHODS SUMMARY

Cases for stage 1 were identified through clinical genetics centres in the UK and a national study of bilateral breast cancer. Cases in stage 2 were drawn from a population-based study of breast cancer (SEARCH)³². Controls for stages 2 and 3 were drawn from EPIC-Norfolk, a population-based study of diet and cancer³³.

Cases and controls for stage 3 were identified through case-control studies in Europe, North America, South-East Asia and Australia participating in the Breast Cancer Association Consortium (Supplementary Table 2)³⁴.

Genotyping for stages 1 and 2 was conducted using high-density oligonucleotide microarrays. For the main analyses, we excluded samples called on $\leq 80\%$ of SNPs in either stage. We also excluded SNPs that achieved a call rate of $\leq 90\%$ in stage 1 and $\leq 95\%$ in stage 2, and SNPs whose frequency deviated from Hardy–Weinberg equilibrium in controls at $P < 0.00001$. Genotyping for stage 3, and for the fine-scale mapping of the *FGFR2* locus, was conducted using either a 5' nuclease assay (Taqman, Applied Biosystems) or MALDI-TOF mass spectrometry using the Sequenom iPLEX system. For each centre, we excluded any sample called on $\leq 80\%$ of SNPs, and any SNP with a call rate of $\leq 95\%$ or a deviation from Hardy–Weinberg equilibrium in controls at $P < 0.00001$. Tests of association were 1 d.f. Cochran–Armitage tests, stratified for stage, centre and ethnic group (European or Asian). Odds ratios for each SNP were estimated using stratified logistic regression, using the stage 3 data only.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 9 February; accepted 30 April 2007.

Published online 27 May 2007; corrected 28 June 2007 (details online).

1. Collaborative Group on Hormonal Factors in Breast Cancer. Familial breast cancer: Collaborative reanalysis of individual data from 52 epidemiological

- studies including 58 209 women with breast cancer and 101 986 women without the disease. *Lancet* **358**, 1389–1399 (2001).
2. Miki, Y. *et al.* A strong candidate for the breast and ovarian-cancer susceptibility gene BRCA1. *Science* **266**, 66–71 (1994).
 3. Wooster, R. *et al.* Identification of the breast cancer susceptibility gene BRCA2. *Nature* **378**, 789–792 (1995).
 4. Antoniou, A. *et al.* Average risks of breast and ovarian cancer associated with mutations in BRCA1 or BRCA2 detected in case series unselected for family history: A combined analysis of 22 studies. *Am. J. Hum. Genet.* **72**, 1117–1130 (2003).
 5. Smith, P. *et al.* A genome wide linkage search for breast cancer susceptibility genes. *Genes Chromosom. Cancer* **45**, 646–655 (2006).
 6. Pharoah, P. D. P. *et al.* Polygenic susceptibility to breast cancer and implications for prevention. *Nature Genet.* **31**, 33–36 (2002).
 7. Antoniou, A. C., Pharoah, P. D. P., Smith, P. & Easton, D. F. The BOADICEA model of genetic susceptibility to breast and ovarian cancer. *Br. J. Cancer* **91**, 1580–1590 (2004).
 8. Rahman, N. *et al.* PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. *Nature Genet.* **39**, 165–167 (2007).
 9. Thompson, D. *et al.* Cancer risks and mortality in heterozygous ATM mutation carriers. *J. Natl Cancer Inst.* **97**, 813–822 (2005).
 10. Meijers-Heijboer, H. *et al.* Low-penetrance susceptibility to breast cancer due to CHEK2*1100delC in noncarriers of BRCA1 or BRCA2 mutations. *Nature Genet.* **31**, 55–59 (2002).
 11. Erkkö, H. *et al.* A recurrent mutation in PALB2 in Finnish cancer families. *Nature* **446**, 316–319 (2007).
 12. Renwick, A. *et al.* ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. *Nature Genet.* **38**, 873–875 (2006).
 13. Seal, S. *et al.* Truncating mutations in the Fanconi anemia J gene BRIP1 are low-penetrance breast cancer susceptibility alleles. *Nature Genet.* **38**, 1239–1241 (2006).
 14. The CHEK2 Breast Cancer Case-Control Consortium. CHEK2*1100delC and susceptibility to breast cancer: A collaborative analysis involving 10,860 breast cancer cases and 9,065 controls from ten studies. *Am. J. Hum. Genet.* **74**, 1175–1182 (2004).
 15. Cox, A. *et al.* A common coding variant in CASP8 is associated with breast cancer risk. *Nature Genetics* **39**, 352–358 (2007); corrigendum **39**, 688 (2007).
 16. Easton, D. F. How many more breast cancer predisposition genes are there? *Breast Cancer Res.* **1**, 1–4 (1999).
 17. Kruglyak, L. & Nickerson, D. A. Variation is the spice of life. *Nature Genet.* **27**, 234–236 (2001).
 18. Hinds, D. A. *et al.* Whole-genome patterns of common DNA variation in three human populations. *Science* **307**, 1072–1079 (2005).
 19. Satagopan, J. M., Verbel, D. A., Venkatraman, E. S., Offit, K. E. & Begg, C. B. Two-stage designs for gene-disease association studies. *Biometrics* **58**, 163–170 (2002).
 20. Antoniou, A. C. & Easton, D. F. Polygenic inheritance of breast cancer: Implications for design of association studies. *Genet. Epidemiol.* **25**, 190–202 (2003).
 21. Altshuler, D. *et al.* A haplotype map of the human genome. *Nature* **437**, 1299–1320 (2005).
 22. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).
 23. Thomas, D. C., Haile, R. W. & Duggan, D. Recent developments in genomewide association scans: A workshop summary and review. *Am. J. Hum. Genet.* **77**, 337–345 (2005).
 24. Amundadottir, L. T. *et al.* A common variant associated with prostate cancer in European and African populations. *Nature Genet.* **38**, 652–658 (2006).
 25. Yeager, M. *et al.* Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nature Genet.* **39**, 645–649 (2007).
 26. Gudmundsson, J. *et al.* Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nature Genet.* **39**, 631–637 (2007).
 27. Ferretti, V. *et al.* PReMod: a database of genome-wide mammalian cis-regulatory module predictions. *Nucleic Acids Res.* **35**, D122–D126 (2007).
 28. Moffa, A. B., Tannheimer, S. L. & Ethier, S. P. Transforming potential of alternatively spliced variants of fibroblast growth factor receptor 2 in human mammary epithelial cells. *Mol. Cancer Res.* **2**, 643–652 (2004).
 29. Adnane, J. *et al.* Bek and Flg, 2 receptors to members of the Fgf family, are amplified in subsets of human breast cancers. *Oncogene* **6**, 659–663 (1991).
 30. Jang, J. H., Shin, K. H. & Park, J. G. Mutations in fibroblast growth factor receptor 2 and fibroblast growth factor receptor 3 genes associated with human gastric and colorectal cancers. *Cancer Res.* **61**, 3541–3543 (2001).
 31. Greenman, C. *et al.* Patterns of somatic mutation in human cancer genomes. *Nature* **446**, 153–158 (2007).
 32. Lesueur, F. *et al.* Allelic association of the human homologue of the mouse modifier Ptp^{trj} with breast cancer. *Hum. Mol. Genet.* **14**, 2349–2356 (2005).
 33. Day, N. *et al.* EPIC-Norfolk: Study design and characteristics of the cohort. *Br. J. Cancer* **80**, 95–103 (1999).
 34. Breast Cancer Association Consortium. Commonly studied SNPs and breast cancer: Negative results from 12,000 – 32,000 cases and controls from the Breast Cancer Association Consortium. *J. Natl Cancer Inst.* **98**, 1382–1396 (2006).
 35. Hubbard, T. *et al.* The Ensembl genome database project. *Nucleic Acids Res.* **30**, 38–41 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors thank the women who took part in this research, and all the funders and support staff who made this study possible. The principal funding for this study was provided by Cancer Research UK. Detailed acknowledgements are provided in Supplementary Information.

Author Contributions D.F.E., A.M.D., P.D.P.P., D.R.C. and B.A.J.P. designed the study and obtained financial support. D.G.B. and D.R.C. directed the genotyping of stages 1 and 2. D.F.E. and D.T. conducted the statistical analysis. K.A.P. and A.M.D. coordinated the genotyping for stage 3 and the fine-scale mapping of the *FGFR2* and *TNRC9* loci. J.P.S. and J.Z. performed resequencing and analysis of the *FGFR2* locus. K.A.P., S.A., C.S.H., R.B., C.A.H., L.K.K., B.E.H., L.L.M., P.B., S.S., V.G., F.O., C-Y. S., P-E.W. and H-C.W. conducted genotyping for the fine-scale mapping. R.L., J.M., H.F. and K.B.M. provided bioinformatics support. D.E., D.G.E., J.P., O.F., N.J., S.S., M.R.S. and N.R. coordinated the studies used in stage 1. N.W. and N.E.D. coordinated the EPIC study used in stages 1 and 2. The remaining authors coordinated the studies in stage 3 and undertook genotyping in those studies. D.F.E. drafted the manuscript, with substantial contributions from K.A.P., A.M.D., P.D.P.P. and B.A.J.P. All authors contributed to the final paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.F.E. (d.easton@srl.cam.ac.uk).

Author affiliations: ¹CR-UK Genetic Epidemiology Unit, Department of Public Health and Primary Care and, ²Department of Oncology, University of Cambridge, Cambridge CB1 8RN, UK. ³Perlegen Sciences, Inc., 2021 Stierlin Court, Mountain View, California 94043, USA. ⁴Laboratory of Population Genetics, US National Cancer Institute, Bethesda, Maryland 20892, USA. ⁵EPIC, Department of Public Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK. ⁶MRC Dunn Clinical Nutrition Centre, Cambridge CB2 0XY, UK. ⁷Cancer Research UK Cambridge Research Institute, Cambridge CB2 0RE, UK. ⁸Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California 90033, USA. ⁹Epidemiology Program, Cancer Research Center of Hawaii, University of Hawaii, Honolulu, Hawaii 96813, USA. ¹⁰International Agency for Research on Cancer, 150 Cours Albert Thomas, Lyon 69008, France. ¹¹National Cancer Institute, Bangkok 10400, Thailand. ¹²Institute of Biomedical Sciences, Academia Sinica, Taipei 11529, Taiwan. ¹³Wessex Clinical Genetics Service, Princess Anne Hospital, Southampton SO16 5YA, UK. ¹⁴Regional Genetic Service, St Mary's Hospital, Manchester M13 0JH, UK. ¹⁵London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK, and Institute of Cancer Research, Sutton, Surrey SM2 5NG, UK. ¹⁶Breakthrough Breast Cancer Research Centre, London SW3 6JB, UK. ¹⁷Section of Cancer Genetics, Institute of Cancer Research, Sutton, Surrey SM2 5NG, UK. ¹⁸Cancer Genome Project, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK. ¹⁹Queensland Institute of Medical Research, Brisbane, Queensland 4006, Australia. ²⁰Departments of Clinical Biochemistry and ²¹Breast Surgery, Herlev and Bispebjerg University Hospitals, University of Copenhagen, DK-2730 Herlev, Denmark. ²²Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, Maryland 20852, USA. ²³Advanced Technology Center, National Cancer Institute, Gaithersburg, Maryland 20877, USA. ²⁴Cancer Center and M. Sklodowska-Curie Institute of Oncology, Warsaw 02781, Poland. ²⁵Nofer Institute of Occupational Medicine, Lodz 90950, Poland. ²⁶Departments of Obstetrics and Gynecology, and ²⁷Department of Oncology, Helsinki University Central Hospital, Helsinki 00029, Finland. ²⁸Seoul National University College of Medicine, Seoul 151-742, Korea. ²⁹National Cancer Center, Goyang 411-769, Korea. ³⁰Ulsan University College of Medicine, Ulsan 680-749, Korea. ³¹Program in Molecular and Genetic Epidemiology, Harvard School of Public Health, 677 Huntington Ave., Boston, Massachusetts 02115, USA. ³²Channing Laboratory, Brigham and Women's Hospital and Harvard Medical School, 181 Longwood Ave., Boston, Massachusetts 02115, USA. ³³Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm SE-171 77, Sweden. ³⁴Population Genetics, Genome Institute of Singapore, 60 Biopolis Street, Singapore 138672, Republic of Singapore. ³⁵Department of Radiation Oncology and ³⁶Department of Gynecology and Obstetrics, Hannover Medical School, D-30625 Hannover, Germany. ³⁷Department of Surgery and ³⁸Department of Medical Decision Making and ³⁹Departments of Human Genetics and Pathology, Leiden University Medical Center, Albinusdreef 2, 2333 ZA Leiden, the Netherlands. ⁴⁰Family Cancer Clinic, Department of Medical Oncology, Erasmus MC-Daniel den Hoed Cancer Center, Groene Hilledijk 301, 3075 EA Rotterdam, the Netherlands. ⁴¹Radiation Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, DHHS, Bethesda, Maryland 20892, USA. ⁴²Environmental Health Sciences, University of Minnesota, Minneapolis, Minnesota 55455, USA. ⁴³Institute for Cancer Studies and ⁴⁴Academic Unit of Surgical Oncology, Sheffield University Medical School, Sheffield S10 2RX, UK. ⁴⁵Mayo Clinic College of Medicine, Rochester, Minnesota 55905, USA. ⁴⁶VU University Medical Center, 1007 MB Amsterdam, the Netherlands. ⁴⁷Department of Clinical Genetics and ⁴⁸Internal Medicine, Erasmus University, Rotterdam NL-3015-GE, the Netherlands. ⁴⁹Spanish National Cancer Centre (CNIO), Madrid E-28029, Spain. ⁵⁰Centre for Molecular, Environmental, Genetic and Analytical Epidemiology, University of Melbourne, Carlton, Victoria 3053, Australia. ⁵¹Department of Preventive and Social Medicine, University of Otago, Dunedin 9001, New Zealand. ⁵²Cancer Epidemiology Centre, Cancer Council Victoria, Carlton, Victoria 3053, Australia. ⁵³Genetic Epidemiology

Laboratory, Department of Pathology, University of Melbourne, Parkville, Victoria 3052, Australia. ⁵⁴Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, 70376 Stuttgart and University of Tuebingen, 72074 Tuebingen, Germany. ⁵⁵Deutsches Krebsforschungszentrum, Heidelberg 69120, Germany. ⁵⁶Evangelische Kliniken Bonn gGmbH Johanniter Krankenhaus, 53113 Bonn, Germany. ⁵⁷Peter MacCallum Cancer Centre, Melbourne, Victoria 3002, Australia. ⁵⁸Institute of Clinical Medicine, Pathology and Forensic Medicine, University of Kuopio, Kuopio FIN-70210, Finland. ⁵⁹Departments of Oncology and Pathology, University Hospital of Kuopio, Kuopio FIN-70211, Finland. ⁶⁰Department of Oncology, Vaasa Central Hospital, Vaasa 65130, Finland.

The SEARCH collaborators Craig Luccarini¹, Don Conroy¹, Mitul Shah¹, Hannah Munday¹, Clare Jordan¹, Barbara Perkins¹, Judy West¹, Karen Redman¹ & Kristy Driver¹. **kConFab** Morteza Aghmehseh², David Amor³, Lesley Andrews⁴, Yoland Antill⁵, Jane Armes⁶, Shane Armitage⁷, Leanne Arnold⁷, Rosemary Balleine⁸, Glenn Begley⁹, John Beilby¹⁰, Ian Bennett¹¹, Barbara Bennett⁴, Geoffrey Bery¹², Anneke Blackburn¹³, Meagan Brennan¹⁴, Melissa Brown¹⁵, Michael Buckley¹⁶, Jo Burke¹⁷, Phyllis Butow¹⁸, Keith Byron¹⁹, David Callen²⁰, Ian Campbell²¹, Georgia Chenevix-Trench²², Christine Clarke²³, Alison Colley²⁴, Dick Cotton²⁵, Jisheng Cui²⁶, Bronwyn Culling²⁷, Margaret Cummings²⁸, Sarah-Jane Dawson⁵, Joanne Dixon²⁹, Alexander Dobrovic³⁰, Tracy Dudding³¹, Ted Edkins³², Maurice Eisenbruch³³, Gelareh Farshid³⁴, Susan Fawcett³⁵, Michael Field³⁶, Frank Firgaira³⁷, Jean Fleming³⁸, John Forbes³⁹, Michael Friedlander⁴⁰, Clara Gaff⁴¹, Mac Gardner⁴¹, Mike Gattas⁴², Peter George⁴³, Graham Giles⁴⁴, Grantley Gill⁴⁵, Jack Goldblatt⁴⁶, Sian Greening⁴⁷, Scott Grist³⁷, Eric Haan⁴⁸, Marion Harris⁴⁹, Stewart Hart⁵⁰, Nick Hayward²², John Hopper⁵¹, Evelyn Humphrey¹⁷, Mark Jenkins⁵², Alison Jones⁷, Rick Keeford⁵³, Judy Kirk⁵⁴, James Kollias⁵⁵, Sergey Kovalenko⁵⁶, Sunil Lakhani⁵⁷, Jennifer Leary⁵⁴, Jacqueline Lim⁵⁸, Geoff Lindeman⁵⁹, Lara Lipton⁶⁰, Liz Lobb⁶¹, Mariette Maclurcan⁶², Graham Mann²³, Deborah Marsh⁶³, Margaret McCredie⁶⁴, Michael McKay⁴⁹, Sue Anne McLachlan⁶⁵, Bettina Meiser⁴, Roger Milne²⁶, Gillian Mitchell⁴⁹, Beth Newman⁶⁶, Imelda O'Loughlin⁶⁷, Richard Osborne⁵¹, Lester Peters⁶⁸, Kelly Phillips⁵, Melanie Price⁶², Jeanne Reeve⁶⁹, Tony Reeve⁷⁰, Robert Richards⁷¹, Gina Rinehart⁷², Bridget Robinson⁷³, Barney Rudzki⁷⁴, Elizabeth Salisbury⁷⁵, Joe Sambrook²¹, Christobel Saunders⁷⁶, Clare Scott⁷⁷, Elizabeth Scott⁷⁷, Rodney Scott³¹, Ram Seshadri³⁷, Andrew Shelling⁷⁸, Melissa Southey²⁶, Amanda Spurdle²², Graeme Suthers⁴⁸, Donna Taylor⁷⁹, Christopher Tennant⁵⁸, Heather Thorne²¹, Sharron Townshend⁴⁶, Kathy Tucker⁴, Janet Tyler⁴, Deon Venter⁸⁰, Jane Visvader⁸¹, Ian Walpole⁴⁶, Robin Ward⁸², Paul Waring³⁰, Bev Warner⁸³, Graham Warren⁶⁷, Elizabeth Watson⁶⁷, Rachael Williams⁸⁴, Judy Wilson⁸⁵, Ingrid Winship⁶⁹ & Mary Ann Young⁴⁹. **AOCS Management Group** David Bowtell⁸⁶, Adele Green²², Anna deFazio⁸⁷, Georgia Chenevix-Trench²², Dorota Gertig⁵¹ & Penny Webb²².

Consortia affiliations: ¹Department of Oncology, University of Cambridge, Cambridge CB1 8RN, UK. ²Oncology Research Centre, Prince of Wales Hospital, Randwick, New South Wales 2031, Australia. ³Genetic Health Services Victoria, Royal Children's Hospital, Melbourne, Victoria 3050, Australia. ⁴Hereditary Cancer Clinic, Prince of Wales Hospital, Randwick, New South Wales 2031, Australia. ⁵Department of Haematology and Medical Oncology, Peter MacCallum Cancer Centre, St Andrews Place, East Melbourne, Victoria 3002, Australia. ⁶Anatomical Pathology, Royal Women's Hospital, Carlton, Victoria 3053, Australia. ⁷Molecular Genetics Laboratory, Royal Brisbane and Women's Hospital, Herston, Queensland 4029, Australia. ⁸Departments of Translational and Medical Oncology, Westmead Hospital, Westmead, New South Wales 2145, Australia. ⁹Cancer Biology Laboratory, TVW Institute for Child Health Research, Subiaco, Western Australia 6008, Australia. ¹⁰Pathology Centre, Queen Elizabeth Medical Centre, Nedlands, Western Australia 6009, Australia. ¹¹Silverton Place, 101 Wickham Terrace, Brisbane, Queensland 4000, Australia. ¹²Department of Public Health and Community Medicine, University of Sydney, Sydney, New South Wales 2006, Australia. ¹³John Curtin School of Medical Research, Australian National University, PO Box 334, Canberra, Australian Capital Territory 2601, Australia. ¹⁴NSW Breast Cancer Institute, PO Box 143, Westmead, New South Wales 2145, Australia. ¹⁵Department of Biochemistry, University of Queensland, St. Lucia, Queensland 4072, USA. ¹⁶Molecular and Cytogenetics Unit, Prince of Wales Hospital, Randwick, New South Wales 2031, Australia. ¹⁷Royal Hobart Hospital, GPO Box 1061L, Hobart, Tasmania 7001, Australia. ¹⁸Medical Psychology Unit, Royal Prince Alfred Hospital, Camperdown, New South Wales 2204, Australia. ¹⁹Australian Genome Research Facility, Walter & Eliza Hall Medical Research Institute, Royal Melbourne Hospital, Parkville, Victoria 3050, Australia. ²⁰Dame Roma Mitchell Cancer Research Laboratories, University of Adelaide/Hanson Institute, PO Box 14, Rundle Mall, South Australia 5000, Australia. ²¹Peter MacCallum Cancer Centre, St Andrew's Place, East Melbourne, Victoria 3002, Australia. ²²Queensland Institute of Medical Research, Herston, Queensland 4006, Australia. ²³Westmead Institute for Cancer Research, University of Sydney, Westmead Hospital, Westmead, New South Wales 2145, Australia. ²⁴Department of Clinical Genetics, Liverpool Health Service, PO Box 103, Liverpool, New South Wales 2170, Australia. ²⁵Mutation Research Centre, St Vincent's Hospital, Victoria Parade, Fitzroy, Victoria 3065, Australia. ²⁶Centre for Genetic Epidemiology, The University of Melbourne, Level 2 723 Swanston Street, Carlton, Victoria 3053, Australia. ²⁷Molecular and Clinical Genetics, Level 1 Building 65, Royal Prince Alfred Hospital, Camperdown, New South Wales 2050, Australia. ²⁸Department of Pathology, University of Queensland Medical School, Herston, New South Wales 4006, Australia. ²⁹Central Regional Genetic Services,

Wellington Hospital, Private bag 7902, Wellington South 6039, New Zealand. ³⁰Molecular Department of Pathology, Peter MacCallum Cancer Centre, St Andrew's Place, East Melbourne, Victoria 3002, Australia. ³¹Hunter Genetics, Hunter Area Health Service, Waratah, New South Wales 2310, Australia. ³²Clinical Chemistry, Princess Margaret Hospital for Children, Box D184, Perth, Western Australia 6001, Australia. ³³Department of Multicultural Health, University of Sydney, New South Wales 2052, Australia. ³⁴Tissue Pathology, Institute of Medical & Veterinary Science, Adelaide, South Australia 5000, Australia. ³⁵Family Cancer Clinic, Monash Medical Centre, Clayton, Victoria 3168, Australia. ³⁶Faculty of Medicine, Royal North Shore Hospital, Vindin House, St Leonards, New South Wales 2065, Australia. ³⁷Department of Haematology, Flinders Medical Centre, Bedford Park, South Australia 5042, Australia. ³⁸Esikit Institute of Cell & Molecular Therapies, School of Biomolecular and Biomedical Sciences, Griffith University, Nathan, Queensland 4111, Australia. ³⁹Surgical Oncology, University of Newcastle, Newcastle Mater Hospital, Waratah, New South Wales 2298, Australia. ⁴⁰Department of Medical Oncology, Prince of Wales Hospital, Randwick, New South Wales 2031, Australia. ⁴¹Victorian Clinical Genetics Service, Royal Melbourne Hospital, Parkville, Victoria 3052, Australia. ⁴²Queensland Clinical Genetic Service, Royal Children's Hospital, Bramston Terrace, Herston, Queensland 4020, Australia. ⁴³Clinical Biochemistry Unit, Canterbury Health Labs, PO Box 151, Christchurch 8140, New Zealand. ⁴⁴Cancer Epidemiology Centre, The Cancer Council Victoria, 1 Rathdowne Street, Carlton, Victoria 3053, Australia. ⁴⁵Department of Surgery, Royal Adelaide Hospital, Adelaide, South Australia 5000, Australia. ⁴⁶Genetic Services of WA, King Edward Memorial Hospital, 374 Bagot Road, Subiaco, Western Australia 6008, Australia. ⁴⁷Wollongong Hereditary Cancer Clinic, Wollongong Public Hospital, Private Mail Bag 8808, South Coast Mail Centre, New South Wales 2521, Australia. ⁴⁸Department of Medical Genetics, Women's and Children's Hospital, North Adelaide, South Australia 5006, Australia. ⁴⁹Familial Cancer Clinic, Peter MacCallum Cancer Centre, St Andrew's Place, East Melbourne, Victoria 3002, Australia. ⁵⁰Breast and Ovarian Cancer Genetics, Monash Medical Centre, 871 Centre Road, Bentleigh East, Victoria 3165, Australia. ⁵¹Centre for Molecular Environmental, Genetic & Analytic Epidemiology, University of Melbourne, Melbourne, Victoria 3010, Australia. ⁵²School of Population Health, The University of Melbourne, 723 Swanston Street, Carlton, Victoria 3053, Australia. ⁵³Medical Oncology, Westmead Hospital, Westmead, New South Wales 2145, Australia. ⁵⁴Familial Cancer Service, Department of Medicine, Westmead Hospital, Westmead, New South Wales 2145, Australia. ⁵⁵Breast Endocrine and Surgical Unit, Royal Adelaide Hospital, North Terrace, South Australia 5000, Australia. ⁵⁶Molecular Pathology Department, Southern Cross Pathology, Monash Medical Centre, Clayton, Victoria 3168, Australia. ⁵⁷Molecular and Cellular Pathology, The University of Queensland, Herston, Queensland 4006, Australia. ⁵⁸Department of Psychological Medicine, Royal North Shore Hospital, St Leonards, New South Wales 2065, Australia. ⁵⁹Breast Cancer Laboratory, Walter and Eliza Hall Institute, PO Royal Melbourne Hospital, Parkville, Victoria 3050, Australia. ⁶⁰Medical Oncology and Clinical Haematology Unit, Western Hospital, Footscray, Victoria 3011, Australia. ⁶¹WA Centre for Cancer, Edith Cowan University, Churchlands, Western Australia 6018, Australia. ⁶²Department of Psychological Medicine, University of Sydney, New South Wales 2006, Australia. ⁶³Kolling Institute of Medical Research, Royal North Shore Hospital, St Leonards, New South Wales 2065, Australia. ⁶⁴Cancer Epidemiology Research Unit, NSW Cancer Council, 153 Dowling Street, Woolloomooloo, New South Wales 2011, Australia. ⁶⁵Department of Oncology, St Vincent's Hospital, 41 Victoria Parade, Fitzroy, Victoria 3065, Australia. ⁶⁶School of Public Health, Queensland University of Technology, Victoria Park, Kelvin Grove, Queensland 4059, Australia. ⁶⁷St Vincent's Breast Clinic, PO Box 4751, Toowoomba, Queensland 4350, Australia. ⁶⁸Radiation Oncology, Peter MacCallum Cancer Centre, St Andrew's Place, East Melbourne, Victoria 3002, Australia. ⁶⁹Genetic Services, Auckland Hospital, Private Bag 92024, Auckland 1142, New Zealand. ⁷⁰Cancer Genetics Laboratory, University of Otago, PO Box 56, Dunedin 9054, New Zealand. ⁷¹Department of Cytogenetics and Molecular Genetics, Women and Children's Hospital, Adelaide, South Australia 5006, Australia. ⁷²Hancock Family Breast Cancer Foundation, PO Locked Bag 2, West Perth, Western Australia 6005, Australia. ⁷³Oncology Service, Christchurch Hospital, Private Bag 4710, Christchurch 8140, New Zealand. ⁷⁴Molecular Pathology Institute of Medical and Veterinary Science, Frome Road, Adelaide, South Australia 5000, Australia. ⁷⁵Section of Cytology, Institute of Clinical Pathology and Medical Research, Westmead Hospital, Westmead, New South Wales 2145, Australia. ⁷⁶School of Surgery and Pathology, QE11 Medical Centre, M block 2nd Floor, Nedlands, Western Australia 6907, Australia. ⁷⁷South View Clinic, Suite 13, Level 3 South Street, Kogarah, New South Wales 2217, Australia. ⁷⁸Department of Obstetrics and Gynaecology, University of Auckland, Private Bag 92019, Auckland 1142, New Zealand. ⁷⁹Department of Radiology, Royal Perth Hospital, Box X2213, Perth 6011, Western Australia, Australia. ⁸⁰Murdoch Institute, Royal Children's Hospital, Parkville, Victoria 3050, Australia. ⁸¹Molecular Genetics of Cancer Division, Walter & Eliza Hall Medical Research Institute, Royal Melbourne Hospital, Parkville, Victoria 3050, Australia. ⁸²Department of Medical Oncology, St Vincents Hospital, Darlinghurst, New South Wales 2010, Australia. ⁸³Cabrini Hospital, 183 Wattletree Road, Malvern, Victoria 3144, Australia. ⁸⁴Family Cancer Clinic, St Vincent's Hospital, Darlinghurst, New South Wales 2010, Australia. ⁸⁵Medical Psychology Research Unit, Royal North Shore Hospital, St Leonards, New South Wales 2065, Australia. ⁸⁶Cancer Genomics & Biochemistry Laboratory, Peter MacCallum Cancer Centre, St Andrew's Place, East Melbourne, Victoria 3002, Australia. ⁸⁷Obstetrics & Gynaecology, Westmead Hospital, University of Sydney, New South Wales 2006, Australia.

METHODS

Subjects. Cases in stage 1 were identified through clinical genetics centres in Cambridge ($n = 91$), Manchester (96) and Southampton (136), and a national study of bilateral breast cancer (85). Cases were women diagnosed with invasive breast cancer under the age of 60 years who had a family history score of at least 2, where the score was computed as the total number of first-degree relatives plus half the number of second-degree relatives affected with breast cancer. The score for women with bilateral breast cancer was increased by 1, so that women were eligible if they were diagnosed with bilateral breast cancer and had one affected first-degree relative. Cases known to carry a *BRCA1* or *BRCA2* mutation were excluded. Controls were selected from the EPIC-Norfolk study, a population-based cohort study of diet and cancer based in Norfolk, East Anglia, UK³³. Controls were chosen to be women aged over 50 years and free of cancer at the time of entry. Genotyping was attempted on 408 cases, plus 32 duplicate case samples, and 400 controls. For the analysis in Table 1, 54 samples with genotype call rates <80% were excluded, so the final analyses were based on 390 cases and 364 controls. The minimum genotype call rate for the remaining samples was 89%. The overall genotype discordance rate between duplicate samples in stage 1 was 0.01%.

For stage 2, invasive breast cancer cases were drawn from SEARCH, a population-based study of cancer in East Anglia³². Controls were women selected from the EPIC-Norfolk study, as previously described³³. Eighty-eight subjects who were also genotyped in stage 1, and 35 controls who subsequently developed breast cancer and were also in the case series, were excluded from the analysis, leaving 3,990 breast cancer cases and 3,916 controls, plus five duplicates. The overall rate of discordance of genotypes between duplicate samples in stage 2 was 0.008%.

Twenty-one additional studies were included in stage 3 (see Supplementary Table 2). These studies participated through the Breast Cancer Association Consortium, an ongoing collaboration among investigators conducting case-control association studies in breast cancer^{15,33}. All studies provided information on disease status (invasive breast cancer, carcinoma *in situ* or control), age at diagnosis/observation, ethnic group, first-degree family history of breast cancer and bilaterality of breast cancer. One further study (Breast Cancer Study of Taiwan) was included in the fine-scale mapping of the *FGFR2* locus.

Genotyping. For stage 1, genotyping was performed on 200 ng DNA that was first subjected to whole genome amplification using Multiple Displacement Amplification (MDA)³⁶. Samples were then genotyped for a set of 266,732 SNPs using high-density oligonucleotide, photolithographic microarrays at Perlegen Sciences. For stage 2, genotyping was performed using 2.5 µg genomic DNA. These samples were genotyped for a set of 13,023 SNPs selected on the basis of the stage 1 results, using a custom designed oligonucleotide array. For both stages, each SNP was interrogated by 24 25-mer oligonucleotide probes synthesized by photolithography on a glass substrate. The 24 features comprise 4 sets of 6 features interrogating the neighbourhoods of SNP reference and alternative alleles on forward and reference strands. Each allele and strand is represented by five offsets: -2, -1, 0, 1 and 2 indicating the position of the SNP within the 25-mer, with zero being at the thirteenth base. At offset 0 a quartet was tiled, which included the perfect match to reference and alternative SNP alleles, and the two remaining nucleotides as mismatch probes. When possible, the mismatch features were selected as a purine nucleotide substitution for a purine perfect match nucleotide and a pyrimidine nucleotide substitution for a pyrimidine perfect match nucleotide. Thus, each strand and allele tiling consisted of 6 features comprising five perfect match probes and one mismatch.

Individual genotypes were determined by clustering all SNP scans in the two-dimensional space defined by reference and alternative trimmed mean intensities, corrected for background. Allele frequencies were approximated using the intensities collected from the high-density oligonucleotide arrays. An SNP's allele frequency, p , was estimated as the ratio of the relative amount of the DNA with reference allele to the total amount of DNA. The \hat{p} value was computed from the trimmed mean intensities of perfect match features, after subtracting a measure of background computed from trimmed means of intensities of mismatch features. The trimmed mean disregarded the highest and the lowest intensity from the five perfect match intensities before computing the arithmetic mean. For the mismatch features, the trimmed mean is the individual intensity of the specified mismatch feature.

The genotype clustering procedure was an iterative algorithm developed as a combination of K-means and constrained multiple linear regressions. The K-means at each step re-evaluated the cluster membership representing distinct diploid genotypes. The multiple linear regressions minimized the variance in \hat{p} within each cluster while optimizing the regression lines' common intersect. The common intersect defined a measure of common background that was used to adjust the allele frequencies for the next step of K-means. The K-means and multiple linear regression steps were iterated until the cluster membership and

background estimates converged. The best number of clusters was selected by maximizing the total likelihood over the possible cluster counts of 1, 2 and 3 (representing the combinations of the three possible diploid genotypes). The total likelihood was composed of data likelihood and model likelihood. The data likelihood was determined using a normal mixture model for the distribution of \hat{p} around the cluster means. The model likelihood was calculated using a prior distribution of expected cluster positions, resulting in optimal \hat{p} positions of 0.8 for the homozygous reference cluster, 0.5 for the heterozygous cluster and 0.2 for the homozygous alternative cluster.

A genotyping quality metric was compiled for each genotype from 15 input metrics that described the quality of the SNP and the genotype. The genotyping quality metric correlated with a probability of having a discordant call between the Perlegen platform and outside genotyping platforms (that is, non-Perlegen HapMap project genotypes). A system of 10 bootstrap aggregated regression trees was trained using an independent data set of concordance data between Perlegen genotypes and HapMap project genotypes. The trained predictor was then used to predict the genotyping quality for each of the genotypes in this data set. Genotypes with quality scores of less than 7 were discarded. Data were analysed for 227,876 SNPs in stage 1 and 12,026 (of 13,023 selected) in stage 2, for which the call rate was >80%.

The 12,711 SNPs for stage 2 were primarily selected on the basis of a 1 d.f. Cochran-Armitage trend test (11,809, all with $P < 0.052$). We also included 826 SNPs with $P < 0.01$ testing for the difference in frequency of either homozygote between cases and controls (that is, assuming either a dominant or recessive model) and 76 SNPs that achieved $P < 0.01$ on a Cochran-Armitage test, weighting individuals by their family history score as above.

For the main analyses, we discarded SNPs with a call rate <90% in stage 1 and 95% in stage 2, and SNPs with a deviation from Hardy-Weinberg equilibrium significant at $P < 0.00001$ in either stage, leaving 205,586 SNPs in stage 1 and 10,621 SNPs in stage 2.

The 30 SNPs included in the stage 3 analyses were initially selected on the basis of a combined analysis of stage 1 and stage 2. We included all SNPs achieving a combined $P < 0.00002$ (based on either the Cochran-Armitage or 2 d.f. test, see below). Following re-evaluation of the stage 2 genotyping by 5' nuclease assay (Taqman, Applied Biosystems) using the ABI PRISM 7900HT (Applied Biosystems), and exclusion of some samples, 16 of these SNPs were significant at $P < 0.00002$ and 24 at $P < 0.0002$ (Supplementary Table 3). One additional SNP, rs3803662, was added as a result of fine-scale mapping of the *TNRC9/LOC643714* locus.

The 31 stage 3 SNPs were genotyped in 22 studies (including cases and controls from SEARCH not used in stage 2, together with 21 other studies). For 18 of the studies, genotyping was performed by 5' nuclease assay (Taqman) using the ABI PRISM 7900HT or 7500 Sequence Detection Systems according to manufacturer's instructions. Primers and probes were supplied directly by Applied Biosystems (<http://www.appliedbiosystems.com/>) as Assays-by-Design. All assays were carried out in 384-well or 96-well format, with each plate including negative controls (with no DNA). Duplicate genotypes were provided for at least 2% of samples in each study. For three studies, SNPs were genotyped using matrix assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF MS) for the determination of allele-specific primer extension products using Sequenom's MassARRAY system and iPLEX technology. The design of oligonucleotides was carried out according to the guidelines of Sequenom and performed using MassARRAY Assay Design software (version 1.0). Multiplex PCR amplification of amplicons containing SNPs of interest was performed using Qiagen HotStart Taq Polymerase on a Perkin Elmer GeneAmp 2400 thermal cycler (MJ Research) with 5 ng genomic DNA. Primer extension reactions were carried out according to manufacturer's instructions for iPLEX chemistry. Assay data were analysed using Sequenom TYPER software (version 3.0). One study used both the Taqman and MALDI-TOF MS approaches. The SNPs genotyped in stage 3 were also regentyped in the stage 2 samples using Taqman; these genotype calls were used in the overall analyses (Table 2, Supplementary Table 3, and Fig. 2).

We eliminated any sample that could not be scored on 20% of the SNPs attempted. We also removed data for any centre/SNP combination for which the call rate was less than 90%. In any instances where the call rate was 90–95%, the clustering of genotype calls was re-evaluated by an independent observer to determine whether the clustering was sufficiently clear for inclusion. We also eliminated all the data for a given SNP/centre where the reproducibility in duplicate samples was <97%, or where there was marked deviation from Hardy-Weinberg equilibrium in the controls ($P < 0.00001$).

Fine-scale mapping of *FGFR2*. Initial tagging of the associated region was done by identifying all SNPs with an m.a.f. > 5% in the HapMap CEPH/CEU set (Utah residents with ancestry from northern and western Europe). We then selected 7 SNPs (in addition to rs2981582) that tagged these variants with a

pairwise $r^2 > 0.8$, using the program Tagger (<http://www.broad.mit.edu/mpg/tagger/>)³⁷. To identify additional common variants within the 32.5 kb region of linkage around the associated SNP, we resequenced 45 lymphocyte DNA samples from a subset of European subjects also genotyped by HapMap and other publicly available data sets. Seventy overlapping PCR amplicons were designed from positions 123317613 to 123348192 of chromosome 10 (average amplicon size 650 bp, 160 bp overlap). M13-tagged PCR products were bidirectionally sequenced using Big Dye 3.0 (Applied Biosystems) and processed using automated trace analysis through the Cancer Genome Workbench (cgwb.nci.nih.gov). Eighty-six per cent of the nucleotides across the region could be scored for polymorphisms in at least 80% of subjects. This set gave a $>97\%$ probability of detecting a variant with an m.a.f. $> 5\%$. One hundred and seventeen variants were identified, including 27 present in dbSNP but without individual genotype information in European subjects, and an additional 46 not in dbSNP. Individual genotype information was then compared and merged with publicly available genotypes from Caucasian subjects (HapMap release 21 for 60 CEU parents, 22 European subjects from the Environmental Genome Project (EGP) resequencing effort (<http://egp.gs.washington.edu/data/fgr2/>), and 24 European subjects from Perlegen (retrieved through <http://gvs.gs.washington.edu/GVS>)). There were 2 discrepancies among 389 genotype calls among subjects in common between our resequencing effort and EGP or Perlegen data, and 10 out of 926 compared to HapMap genotypes.

On the basis of these data, we identified 28 SNPs correlated with rs2981582 with $r^2 > 0.6$. We then attempted to genotype these 28 SNPs, plus rs2981582, in a subset of 80 controls from SEARCH and 84 controls from the Seoul Breast Cancer Study. Twenty-two of the variants were genotyped using Taqman. Four further variants (rs34032268, rs2912778, rs2912781 and rs7895676), which were not amenable to Taqman, were genotyped by Pyrosequencing (Biotage; <http://www.biotagebio.com/>). Assays were designed using Pyrosequencing Assay Design Software 1.0. The remaining 2 SNPs (rs35393331 and rs33971856) could not be genotyped using either technology and were excluded from further analyses. We cannot therefore comment on their likelihood of being the causal variant. Using these data, we selected tagging sets of 11 SNPs for UK subjects and 14 SNPs for Korean subjects (including rs2981582), such that each of the remaining variants was correlated with a tagging SNP with $r^2 > 0.95$ in the UK study or $r^2 > 0.86$ in the Korean study. After genotyping the 11 tag SNPs in SEARCH, two of these SNPs (rs4752569 and rs35012336) showed strong evidence against being the causative variant and were not considered further. The remaining 12 tag SNPs from the Korean subset were then genotyped in the samples from the IARC-Thai Breast Cancer Study, the Breast Cancer Study in Taiwan and the Multi-Ethnic Cohort (MEC), by Taqman.

Statistical methods. The primary test used for each SNP was a Cochran-Armitage 1 d.f. score test for association between disease status and allele dose. In the combined analysis, we performed a stratified Cochran-Armitage test. Stage 1 was given a weight of 4 in this analysis (corresponding to a weight of 2 in the score statistic), to allow for the expected greater effect size given the inclusion of cases with a family history. In the stage 3 analyses, each study was treated as a separate stratum, except for the MEC, in which the European American and Japanese American subgroups were treated as separate strata. For all studies except the MEC, individuals from a minor ethnic group for that study were excluded. Per-allele and genotype-specific odds ratios, and confidence intervals, were estimated using logistic regression, adjusting for the same strata. The summary odds ratios in Fig. 2 are based on the data from the stage 3 studies only, to avoid the bias inherent in estimates from the stage 1 and 2 data for SNPs exhibiting an association (the so called 'winner's curse'). The effects of genotype on family history of breast cancer (first degree yes/no) and bilaterality were examined by treating these variables as outcomes in a stratified Cochran-Armitage test.

To assess the global significance of the SNPs in stage 3, we computed the sum of the χ^2 trend statistics (excluding the 6 SNPs reaching genome-wide significance, plus rs2107425 as it was in LD with rs3817198) over those SNPs (17 of 23) for which the estimated odds ratios in stage 3 were in the same direction as the combined stage 1/stage 2³⁸. Under the null hypothesis of no association, the asymptotic distribution of this statistic is χ^2 with n degrees of freedom, where n has a binomial distribution with parameters 23 and 1/2. The significance of this statistic was then assessed by computing a weighted sum of the tails of the relevant χ^2 distributions.

For the fine-scale mapping of the *FGFR2* locus, we first derived haplotype frequencies using the haplo.stats package in S-plus³⁹, separately for the European and Asian populations, using data from the case-control studies on whom the tag SNPs were typed plus the 164 control individuals on whom all SNPs were typed. These were used to impute genotype probabilities for each identified SNP in each individual. We then used an EM algorithm to fit a logistic regression model assuming that each SNP in turn was the causal variant, allowing for uncertainty

in the genotypes of untyped SNPs, and hence to determine the likelihood that each SNP was the causal variant.

Coverage of the stage 1 tagging set was estimated using HapMap phase II as a reference. We based estimates on 2,116,183 SNPs with an m.a.f. of $>5\%$ in the CEU population. Of the SNPs successfully genotyped in stage 1, 187,663 were also on HapMap. For those SNPs not on HapMap, we identified 'surrogate' SNPs that were in perfect LD based on genotyping of 24 Caucasians by Perlegen Sciences (269,203 SNPs)¹⁸. To estimate coverage, we determined the best pairwise r^2 for each HapMap SNP and each tag SNP or a surrogate SNP, using the HapMap CEU data. This coverage was summarized in terms of the distribution of r^2 by allele frequency in 10 categories.

To estimate the power to detect each of the associations found, we computed the non-centrality parameter for the test statistic at each stage, based on the per-allele relative risk, allele frequency and r^2 . This was used to estimate the power for a given r^2 , based on a simulated trivariate normal distribution for the score statistics after each stage to allow for the correlations in the test statistics. We assumed a cut-off of $P < 0.05$ for stage 1, $P < 0.00002$ for stage 2 and $P < 10^{-7}$ for stage 3 (the first is slightly conservative, as more SNPs than this were actually taken forward). The overall power was obtained by averaging the power estimates for each r^2 over the distribution of r^2 obtained from the HapMap data, applicable to a SNP of that frequency.

The expected number of significant associations after stage 2 (Table 1) was calculated using a bivariate normal distribution for the joint distribution of the (weighted) Cochran-Armitage score statistics after stage 1 and after both stages, using a correlation of 0.525 between the two statistics (reflecting the weighted sizes of the two studies). These calculations were based on the 205,586 SNPs reaching the required quality control in stage 1. Of these, 11,313 reached a $P < 0.05$, of which 7,405 (65.5%) were successfully genotyped to the required quality control in stage 2. Thus the expected number reaching a given significance level with good quality control was calculated from the total number expected to reach this level $\times 65.5\%$. We adjusted the variances of the test statistics, separately for stages 1 and 2, using the genomic control method²². The adjustment factor, λ , was estimated from the median of the smallest 90% of the test statistics for SNPs typed in that stage, divided by the predicted median for the smallest 90% of a sample of χ^2_1 distributions (that is, the 45% percentile of a χ^2_1 distribution, 0.375).

36. Dean, F. B. *et al.* Comprehensive human genome amplification using multiple displacement amplification. *Proc. Natl Acad. Sci. USA* **99**, 5261–5266 (2002).
37. de Bakker, P. I. W. *et al.* Efficiency and power in genetic association studies. *Nature Genet.* **37**, 1217–1223 (2005).
38. Tyrer, J., Pharoah, P. D. P. & Easton, D. F. The admixture maximum likelihood test: A novel experiment-wise test of association between disease and multiple SNPs. *Genet. Epidemiol.* **30**, 636–643 (2006).
39. Schaid, D. J., Rowland, C. M., Tines, D. E., Jacobson, R. M. & Poland, G. A. Score tests for association between traits and haplotypes when linkage phase is ambiguous. *Am. J. Hum. Genet.* **70**, 425–434 (2002).

LETTERS

Chemical complexity in the winds of the oxygen-rich supergiant star VY Canis Majoris

L. M. Ziurys^{1,2,3,4}, S. N. Milam^{1,2,4}, A. J. Apponi^{1,2,4} & N. J. Woolf^{1,2}

The interstellar medium is enriched primarily by matter ejected from old, evolved stars^{1,2}. The outflows from these stars create spherical envelopes, which foster gas-phase chemistry^{3–5}. The chemical complexity in circumstellar shells was originally thought to be dominated by the elemental carbon to oxygen ratio⁶. Observations have suggested that envelopes with more carbon than oxygen have a significantly greater abundance of molecules than their oxygen-rich analogues⁷. Here we report observations of molecules in the oxygen-rich shell of the red supergiant star VY Canis Majoris (VY CMa). A variety of unexpected chemical compounds have been identified, including NaCl, PN, HNC and HCO⁺. From the spectral line profiles, the molecules can be distinguished as arising from three distinct kinematic regions: a spherical outflow, a tightly collimated, blue-shifted expansion, and a directed, red-shifted flow. Certain species (SiO, PN and NaCl) exclusively trace the spherical flow, whereas HNC and sulphur-bearing molecules (amongst others) are selectively created in the two expansions, perhaps arising from shock waves. CO, HCN, CS and HCO⁺ exist in all three components. Despite the oxygen-rich environment, HCN seems to be as abundant as CO. These results suggest that oxygen-rich shells may be as chemically diverse as their carbon counterparts.

Circumstellar envelopes are unique chemical laboratories. Molecules are thought to form near the hot, dense stellar photosphere by a combination of equilibrium and non-equilibrium (that

is, shocks and stellar photodissociation) processes⁸. As the gas flows from the star, molecular abundances ‘freeze-out’ until the outer edge, where the chemistry is further modified by interstellar ultraviolet radiation^{3,4}. Chemical compounds in these shells apparently survive thousands of years as the star evolves to its final stage⁵.

Oxygen is typically more abundant than carbon on the surface of main sequence stars, with a roughly solar ratio of C/O = 0.5 (refs 6, 9). Stars in a certain mass range, however, pass through the asymptotic giant branch (AGB), where they undergo substantial convection. ‘Dredge-up’ events mix carbon from the interior helium-burning shell to the surface such that the C/O ratio becomes >1 (ref. 6), resulting in ‘carbon-rich’ as opposed to ‘oxygen-rich’ envelopes, which vary in chemical complexity. The C-rich envelope of the AGB star IRC + 10216, for example, contains over 60 different chemical compounds, including unusual carbon chain radicals⁵. In contrast, only 10–12 compounds have been identified in the most chemically interesting oxygen-rich evolved stars, such as IK Tau¹⁰.

VY Canis Majoris (VY CMa) is an oxygen-rich¹¹ supergiant (~25 solar masses, 25M_⊙) star with a luminosity of 500,000 times that of the Sun and an effective temperature near 3,000 K (ref. 12). The mass loss rate of VY CMa is substantial (~2 × 10^{−4} M_⊙ yr^{−1}), resulting in a clumpy envelope consisting of arcs, knots and jets on the scale of 10'' (~15,000 AU), evident in Hubble Space Telescope images¹² (Fig. 1). Molecular line observations trace a shell expanding at a velocity of ~40 km s^{−1} (ref. 13). Aperture synthesis maps of CO and maser

Table 1 | Molecular abundances in VY Canis Majoris

Molecule*	Source radius (arcsec)		Abundance relative to H ₂		
	Spherical wind	Red/blue flow	Spherical wind	Red-shifted flow	Blue-shifted flow
CN	6	8.5	2 × 10 ^{−8}	3 × 10 ^{−8}	1 × 10 ^{−8}
CO	6	8.5	5 × 10 ^{−5}	6 × 10 ^{−5}	8 × 10 ^{−5}
CS	0.5	0.7	1 × 10 ^{−7}	4 × 10 ^{−8}	1 × 10 ^{−7}
H ₂ O	0.1		4 × 10 ^{−4} †		
H ₂ S	6	8.5	7 × 10 ^{−8}	2 × 10 ^{−7}	1 × 10 ^{−7}
HCN	3	3.5	8 × 10 ^{−5}	4 × 10 ^{−5}	4 × 10 ^{−5}
HCO⁺	6	8.5	2 × 10 ^{−8}	2 × 10 ^{−8}	2 × 10 ^{−8}
HNC		0.7		2 × 10 ^{−8}	2 × 10 ^{−8}
NaCl	0.25		8 × 10 ^{−9}		
NH ₃				4 × 10 ^{−6} ‡	4 × 10 ^{−6} ‡
NS		8.5		1 × 10 ^{−8} ‡	6 × 10 ^{−9} ‡
OH				maser	maser
PN	0.5		4 × 10 ^{−8}		
SiO	6		~1 × 10 ^{−5}		
SiS	0.5	0.7	7 × 10 ^{−6}	2 × 10 ^{−6}	7 × 10 ^{−7}
SO		8.5		5 × 10 ^{−8}	4 × 10 ^{−8}
SO ₂		8.5		4 × 10 ^{−7}	3 × 10 ^{−7}

Abundances are relative to H₂. Abundances and source sizes were derived by modelling the line profiles, assuming an appropriate geometry and mass loss rate. In almost all cases, at least two transitions were simultaneously fitted to establish abundances. Collisional excitation was assumed. See Supplementary Information for details.

* Bold denotes first time observed towards VY CMa.

† Ref. 28.

‡ Ref. 29.

¹NASA Astrobiology Institute, ²Department of Astronomy/Steward Observatory, ³Arizona Radio Observatory, University of Arizona, 933 North Cherry Avenue, ⁴Department of Chemistry, University of Arizona, 1306 East University Boulevard, Tucson, Arizona 85721, USA.

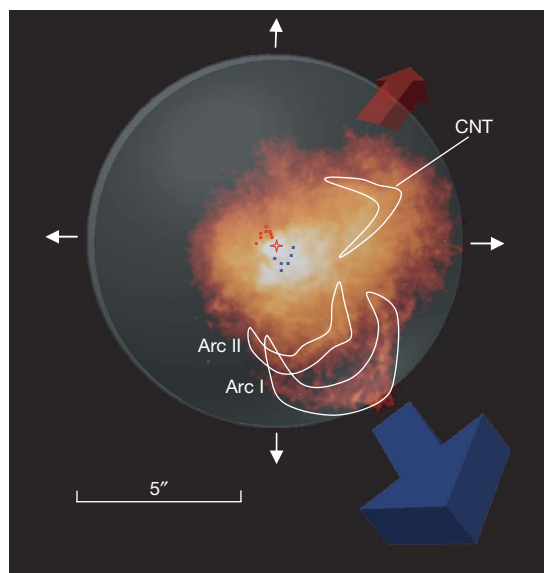


Figure 1 | Model of the molecular outflows in VY CMa, superimposed on the HST infrared image. The spherical wind is represented by the white arrows and the semi-transparent circle, whereas the two directed expansions are indicated by the (large) blue and red arrows. Positions of the red- and blue-shifted OH masers are shown (small rectangles) relative to the central star (red diamond), as well as the locations of the infrared features Arc 1 and Arc 2, located to the southwest, and the Curved Nebulous Tail (CNT), which appears northwest of the star¹². The blue-shifted flow is slightly inclined relative to the line of sight, following the orientation of the masers. The arcs seem to follow the blue-shifted expansion. The red-shifted flow is oriented at a larger angle relative to the line of sight, roughly in the direction of the CNT. These infrared features probably trace shocks associated with the two expansions.

emission from OH, H₂O and SiO indicate structure of the order of an arcsecond, suggesting the presence of a tilted, expanding disk or a bipolar outflow^{14,15}.

Oxygen-rich circumstellar shells have never been studied in equivalent detail to the carbon-rich counterparts. Therefore, we began a spectral survey of VY CMa at wavelengths of 1, 2 and 3 mm (75–270 GHz) using the 12-m telescope and the Sub-millimeter Telescope (SMT) of the Arizona Radio Observatory. In the course of our observations, seven molecules new to this source were discovered (NaCl, HCO⁺, PN, CS, NS, HNC and SiS), bringing to 17 the total number of chemical compounds found in the envelope of VY CMa (Table 1).

The spectral-line profiles varied between species. SiO and PN were found to have roughly triangular line shapes (Fig. 2), tracing a spherical wind concentrated in the plane of the sky, with line widths $\Delta v_{1/2}$ (full-width at half-maximum) near 40 km s⁻¹. NaCl exhibited a much narrower line profile ($\Delta v_{1/2} \approx 14$ km s⁻¹), and arises from the interior region where the maximum outflow velocity has not yet been attained. SO₂ and SO display asymmetric horn-shaped profiles, with distinct blue- and red-shifted components near -7 km s⁻¹ and 42 km s⁻¹, respectively. The horned profile in SO₂ is strikingly similar to that found in the OH 1,612 MHz maser transition¹⁴, as shown in Fig. 2. SO₂, however, exhibits the same line shape in multiple transitions over a wide range of energies, consistent with a Boltzmann distribution with temperature $T \approx 50$ K. Furthermore, the blue-shifted feature in both molecules is almost identical in shape, suggesting a collimated flow close to the line of sight. In

disappears before the terminal outflow velocity is achieved. SO₂ and the OH maser profiles (f and g) are remarkably similar, suggesting that these species are linked chemically and dynamically. The narrow spike that appears in the CO spectrum near 25 km s⁻¹ in LSR velocity arises from background gas. Spectral resolution for all data except the OH line is 1 MHz.

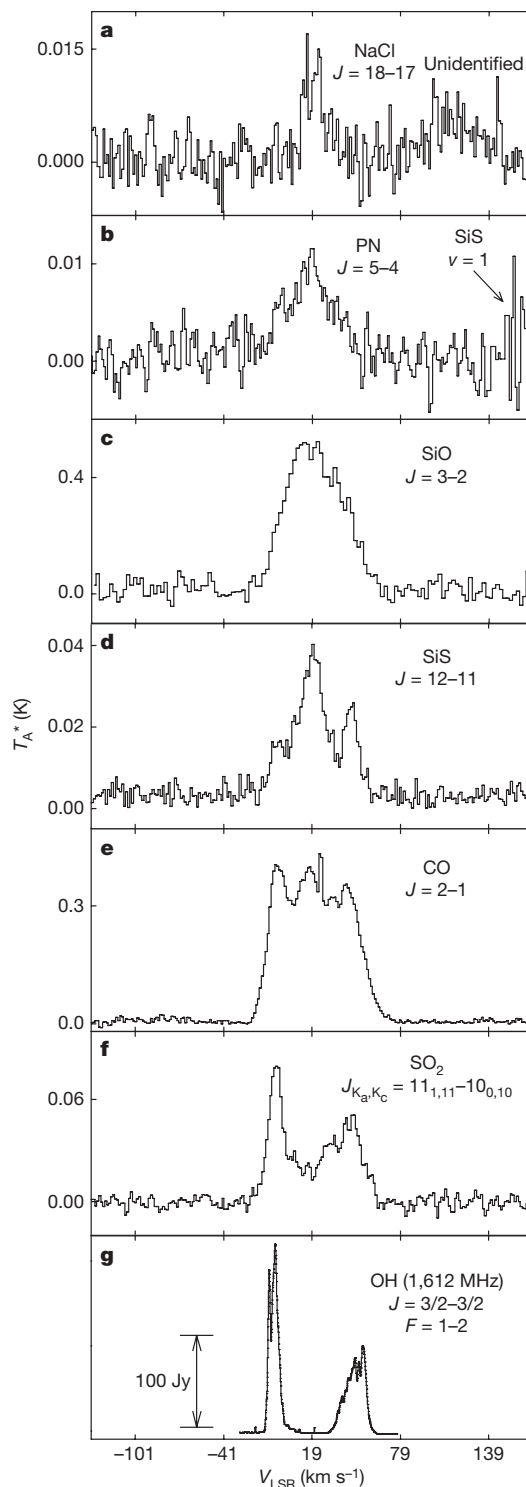


Figure 2 | Sample molecular spectra from VY CMa, measured with the Arizona Radio Observatory's Sub-millimeter Telescope (SMT), showing the variation in line profiles. The receiver used for these observations was an ALMA Band 6 prototype system³⁰, which produced exceptional total system temperatures (single sideband) of $T_{\text{sys}} \approx 120$ –140 K on the sky. Various rotational transitions of molecules measured in the 1–2-mm wavelength region are shown (a–f), as well as the OH 1,612 MHz maser emission from ref. 14 (g). The spectrum in c was obtained with the 12-m telescope. The spectra are plotted in terms of intensity (T_A^* , in K) versus velocity with respect to the local standard of rest (V_{LSR} , in km s⁻¹). The intensity scale for the OH data are in Janskys (Jy). The quantum numbers of the particular rotational transition displayed are given to the right of each spectrum. As these data illustrate, NaCl (a), PN (b) and SiO (c) exclusively trace the spherical wind, whereas the two directed outflows are prominent in SO₂ (f). SiS (d) and CO (e) are found in all three regions. The NaCl line is particularly narrow because this molecule

contrast, the red-shifted emission in SO₂ lies at lower projected expansion velocities than the terminal value indicated by this species itself and the maser emission. This difference indicates that the bulk of the red-shifted material is moving at a substantial angle relative to the line of sight. Many species, including SiS, HCN, H₂S, CO and CS, are present in all three flows. Comparison with previous CO and HCN data suggests that these regions extend to radii 6''–8'' from the central star^{16,17}.

The similarity of the OH and SO₂ line profiles indicate that the blue-shifted outflow is oriented to the southwest, like the masers, and encompasses the infrared dust features Arc 1 and Arc 2 (see Fig. 1). Potassium emission lines from Arc 1 also show a strong, blue-shifted feature with a velocity almost identical to that of the molecular spectra^{18,19}, suggesting that the arcs trace shocks created in the flow. Atomic absorption lines from the 'Curved Nebulous Tail' (CNT) have the same peak velocities as the red-shifted molecular lines. The red-shifted outflow therefore probably lies to the northwest, with the masers on the eastern edge. These two expansions seem to be unrelated; they exhibit different line shapes and molecular rotational temperatures. Large convective cells²⁰, as observed in the supergiant Betelgeuse²¹, could be their origin.

The chemistry in the envelope of VY CMa differs with physical location. NaCl is present only in the spherical flow close to the photosphere, and probably condenses into grains at a radius near 0.25''. NaCl has previously been observed in the carbon-rich shells of two AGB stars^{5,22}, but is more abundant in VY CMa. SiO, in contrast, survives well into the spherical flow, reaching the terminal velocity. The abundance of this species is estimated to be $\sim 10^{-5}$, relative to H₂, in a 12'' region. (Abundances quoted here are relative to H₂, unless stated otherwise: see Supplementary Information.) A large fraction of the elemental silicon is therefore in gas-phase SiO, as opposed to silicate particles. SiS has a narrower line width than its oxide counterpart ($\sim 25 \text{ km s}^{-1}$ versus $\sim 42 \text{ km s}^{-1}$), suggesting that it condenses out or reacts to form other species, well before SiO. Modelling of the observed SiS transitions indicates a source size of 1'', consistent with an early disappearance, and an abundance of 7×10^{-6} (Supplementary Information). Unlike SiO, however, SiS also appears in the red- and blue-shifted expansions. PN exists exclusively in the inner part of the spherical flow with a 1'' source size and an abundance of $\sim 4 \times 10^{-8}$. These observations are the first conclusive identification of PN in circumstellar gas²³.

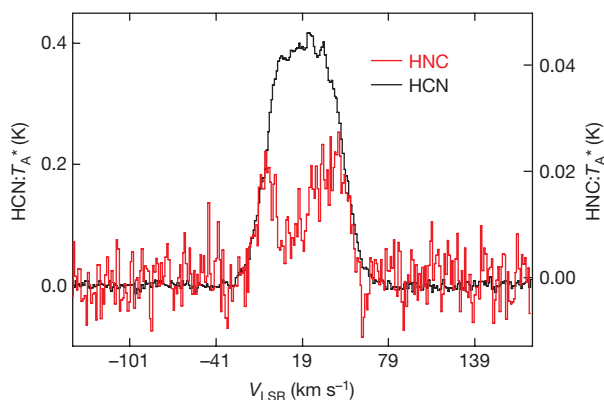


Figure 3 | Comparison of spectra for the metastable isomers HCN and HNC in VY CMa. The $J = 3 \rightarrow 2$ rotational transitions of both species are shown in the figure. The HNC spectrum is displayed in red, whereas that for HCN is plotted in black. HNC shows only the red- and blue-shifted flows, whereas HCN has a contribution from the spherical wind as well. The HNC/HCN ratio is roughly 0.001 for the directed flows, indicating high temperature formation²⁶, perhaps due to shocks. The spectra were measured in the 1-mm wavelength band using the SMT. The scale on the left-hand side of the spectra applies to the HCN data, whereas that on the right is for HNC. Spectral resolution for both lines is 1 MHz.

Six carbon-bearing species are present in VY CMa: CO, HCN, HNC, CN, CS and HCO⁺. Shock and 'mixed' chemistry models have predicted the synthesis of HCN and CS in O-rich environments^{8,17,24}, and these species indeed are observed in several O-rich shells^{17,25}. The models compute an HCN abundance at least a factor of ten less than CO, whereas observations have found HCN/CO < 1/100. In contrast, CO and HCN in VY CMa seem to have comparable abundances in all three flows, roughly 5×10^{-5} (Table 1). The appearance of HCO⁺ is completely unexpected; this molecular ion had not been identified previously in circumstellar gas, and its existence in such material is not predicted by 'mixed' chemistry models^{8,17}. HNC, the metastable isomer of HCN (ref. 26), is also an unusual species for O-rich shells²⁵. This molecule is present exclusively in the directed flows, probably generated by the shock dissociation of HCN as the expansions drive into the spherical wind (Fig. 3).

SO₂, NS and SO predominantly exist in the red- and blue-shifted outflows. The similarity of their line profiles to OH suggests these species are related dynamically. Models have predicted that the creation of OH quickly leads to SO and SO₂, provided there is available sulphur^{13,27}. Sulphur may arise from the dissociation of CS and SiS, and OH is a shock product of H₂O. Nitrogen for NS is probably produced from dissociated HCN. For both SiS and HNC, the red-shifted component seems to have stronger emission, whereas the blue-shifted wing is dominant in SO₂, SO and NS, showing chemical variation and suggesting shock differences in the two expansions.

Molecular production in VY CMa and its associated morphology cannot be explained by the usual spherical models of circumstellar synthesis, even when modified by non-equilibrium or 'mixed' chemistries. Shock waves and grain sputtering are likely to be influencing abundances. The diverse chemistry of VY CMa may be characteristic of red supergiant stars, induced by their complex process of mass loss.

Received 15 March; accepted 4 May 2007.

- Marvel, K. B. No methane here. The HCN puzzle: Searching for CH₃OH and C₂H in oxygen-rich stars. *Astron. J.* **130**, 261–268 (2005).
- Wilson, L. A. Mass loss from cool stars: Impact on the evolution of stars and stellar populations. *Annu. Rev. Astron. Astrophys.* **38**, 573–611 (2000).
- McCabe, E. M., Smith, R. C. & Clegg, R. E. S. Molecular abundances in IRC+10216. *Nature* **281**, 263–266 (1979).
- Glassgold, A. E. Circumstellar photochemistry. *Annu. Rev. Astron. Astrophys.* **34**, 241–278 (1996).
- Ziurys, L. M. The chemistry in circumstellar envelopes of evolved stars: Following the origin of the elements to the origin of life. *Proc. Natl Acad. Sci. USA* **103**, 12274–12279 (2006).
- Iben, I. & Renzini, A. Asymptotic giant branch evolution and beyond. *Annu. Rev. Astron. Astrophys.* **21**, 271–342 (1983).
- Oloffson, H. Molecules in envelopes around AGB-stars. *Astrophys. Space Sci.* **251**, 31–39 (1997).
- Cherchneff, I. A chemical study of the inner winds of asymptotic giant branch stars. *Astron. Astrophys.* **456**, 1001–1012 (2006).
- Anders, E. & Grevesse, N. Abundances of the elements: Meteoritic and solar. *Geochim. Cosmochim. Acta* **53**, 197–214 (1989).
- Duari, D., Cherchneff, I. & Willacy, K. Carbon molecules in the inner wind of the oxygen-rich Mira IK Tauri. *Astron. Astrophys.* **341**, L47–L50 (1999).
- Wallerstein, G. & Gonzalez, G. The spectrum of VY Canis Majoris in 2000 February. *Publ. Astron. Soc. Pacif.* **113**, 954–956 (2001).
- Smith, N. et al. The asymmetric nebula surrounding the extreme red supergiant VY Canis Majoris. *Astron. J.* **121**, 1111–1125 (2001).
- Sahai, R. & Wannier, P. G. SO and SO₂ in mass-loss envelopes of red giants: Probes of nonequilibrium circumstellar chemistry and mass-loss rates. *Astrophys. J.* **394**, 320–339 (1992).
- Bowers, P. F., Johnston, K. J. & Spencer, J. H. Circumstellar envelope structure of late-type stars. *Astrophys. J.* **274**, 733–754 (1983).
- Muller, S. et al. The molecular envelope around the red supergiant VY CMa. *Astrophys. J.* **656**, 1109–1120 (2007).
- Kemper, F. et al. Mass loss and rotational CO emission from asymptotic giant branch stars. *Astron. Astrophys.* **407**, 609–629 (2003).
- Nercessian, E., Guilloteau, S., Omont, A. & Benayoun, J. J. HCN emission and nitrogen-bearing molecules in oxygen-rich circumstellar envelopes. *Astron. Astrophys.* **210**, 225–235 (1989).
- Humphreys, R. M., Davidson, K., Ruch, G. & Wallerstein, G. High-resolution, long-slit spectroscopy of VY Canis Majoris: The evidence for localized high mass loss events. *Astron. J.* **129**, 492–510 (2005).
- Smith, N. Spatially extended K I $\lambda 7699$ emission in the nebula of VY CMa: kinematics and geometry. *Mon. Not. R. Astron. Soc.* **349**, L31–L35 (2004).

20. Schwarzschild, M. On the scale of photospheric convection in red giants and supergiants. *Astrophys. J.* **195**, 137–144 (1975).
21. Lim, J., Carilli, C. L., White, S. M., Beasley, A. J. & Marson, R. G. Large convective cells as the source of Betelgeuse's extended atmosphere. *Nature* **392**, 575–577 (1998).
22. Highberger, J. L., Thomson, K. J., Young, P. A., Arnett, D. & Ziurys, L. M. The salty scrambled egg: Detection of NaCl toward CRL 2688. *Astrophys. J.* **593**, 393–401 (2003).
23. Cernicharo, J., Guelin, M. & Kahane, C. A λ 2 mm molecular line survey of the C-star envelope IRC+10216. *Astron. Astrophys.* **142** (Suppl.), 181–215 (2000).
24. Szczerba, R., Schmidt, M. R. & Pulecka, M. Mixed chemistry phenomenon during late stages of stellar evolution. *Balt. Astron.* **16**, 134–141 (2007).
25. Bujarrabal, V., Fuente, A. & Omont, A. Molecular observations of O- and C-rich circumstellar envelopes. *Astron. Astrophys.* **285**, 247–271 (1994).
26. Schilke, P. et al. A study of HCN, HNC and their isotopomers in OMC-1. *Astron. Astrophys.* **256**, 595–612 (1992).
27. Nejad, L. A. M. & Millar, T. J. Chemical modeling of molecular sources – VI. Carbon-bearing molecules in oxygen-rich circumstellar envelopes. *Mon. Not. R. Astron. Soc.* **230**, 79–86 (1988).
28. Zubko, V., Li, D., Lim, T., Feuchtgruber, H. & Harwit, M. Observations of water vapour outflow from NML Cygnus. *Astrophys. J.* **610**, 427–435 (2004).
29. Monnier, J. D., Danchi, W. C., Hale, D. S., Tuthill, P. G. & Townes, C. H. Mid-infrared interferometry on spectral lines. III. Ammonia and silane around IRC+10216 and VY Canis Majoris. *Astrophys. J.* **543**, 868–879 (2000).
30. Lauria, E. F. et al. *First Astronomical Observations with an ALMA Band 6 (211–275 GHz) Sideband-Separating SIS Mixer-preamp* (ALMA Memo No. 553); (www.alma.nrao.edu/memos/) (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the National Radio Astronomy Observatory for the loan of the ALMA Band 6 mixer system, and A. Lichtenberger and the University of Virginia Microfabrication Laboratory for supplying the mixer junctions. This research is partly supported by the NSF Astronomy and NASA Astrobiology programmes.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to L.M.Z. (lziurys@as.arizona.edu).

LETTERS

Tunable nanowire nonlinear optical probe

Yuri Nakayama^{1,6*}, Peter J. Pauzauskie^{1,5*}, Aleksandra Radenovic^{2,4*}, Robert M. Onorato^{1*}, Richard J. Saykally^{1,4}, Jan Liphardt^{2,3,4} & Peidong Yang^{1,5}

One crucial challenge for subwavelength optics has been the development of a tunable source of coherent laser radiation for use in the physical, information and biological sciences that is stable at room temperature and physiological conditions. Current advanced near-field imaging techniques using fibre-optic scattering probes^{1,2} have already achieved spatial resolution down to the 20-nm range. Recently reported far-field approaches for optical microscopy, including stimulated emission depletion³, structured illumination⁴, and photoactivated localization microscopy⁵, have enabled impressive, theoretically unlimited spatial resolution of fluorescent biomolecular complexes. Previous work with laser tweezers^{6–8} has suggested that optical traps could be used to create novel spatial probes and sensors. Inorganic nanowires have diameters substantially below the wavelength of visible light and have electronic and optical properties^{9,10} that make them ideal for subwavelength laser and imaging technology. Here we report the development of an electrode-free, continuously tunable coherent visible light source compatible with physiological environments, from individual potassium niobate (KNbO₃) nanowires. These wires exhibit efficient second harmonic generation, and act as frequency converters, allowing the local synthesis of a wide range of colours via sum and difference frequency generation. We use this tunable nanometric light source to implement a novel form of subwavelength microscopy, in which an infrared laser is used to optically trap and scan a nanowire over a sample, suggesting a wide range of potential applications in physics, chemistry, materials science and biology.

Nanometre-scale photonics is emerging as a key ingredient in novel sensing and imaging applications, as well as for advanced information technology, cryptography, and signal processing circuits. A versatile and useful nonlinear circuit element for integrated optical networks must be able to double the frequency of light using second harmonic generation (SHG), a second-order nonlinear optical phenomenon. In this process, two photons with the fundamental angular frequency ω_1 are converted through a nonlinear crystal polarization into a single photon ω_2 at twice the fundamental frequency ($\omega_2 = 2\omega_1$). We have recently demonstrated and characterized harmonic generation¹¹, waveguiding, and optically pumped lasing in single nanowires of zinc oxide¹² and gallium nitride¹³. Despite the growing availability of building blocks such as light-emitting diodes^{10,14}, lasers^{13,15,16}, photodetectors¹⁷, and waveguides¹⁸, the field still lacks sufficiently small devices that efficiently generate tunable coherent photons. Here we show that the large second-order susceptibility $\chi^{(2)}$ of KNbO₃ nanowires facilitates the generation of tunable, coherent visible radiation that is sufficient for *in situ* scanning and fluorescence microscopy.

We chose the perovskite oxide KNbO₃ as the nanowire material because of its low toxicity, chemical stability, large effective nonlinear optical coefficients ($d_{\text{eff}} = 10.8\text{--}27\text{ pm V}^{-1}$ at $\lambda = 1,064\text{ nm}$) at room temperature (298 K)¹⁹, large refractive indices ($n = 2.1\text{--}2.5$)²⁰, as well as its transparency in a wide range of wavelengths including the

visible spectral region²¹. Single-crystalline KNbO₃ nanowires were synthesized using a hydrothermal method²², and characterized as orthorhombic phase (*Amm*2) with the growth axis parallel to the [011] direction (Fig. 1a–e): the polar *c* axis²³ is therefore 45° off the nanowire's growth axis.

The SHG response of single KNbO₃ nanowires were characterized first using femtosecond pulses described elsewhere²⁴ and compared with measurements of ZnO nanowires which have been studied previously¹¹. The nanowires were supported on amorphous silica coverslips and aligned such that the growth axis was orthogonal to the pump beam. The maximum SHG signals for both KNbO₃ and ZnO nanowires ($\lambda = 502\text{ nm}$) are shown in Fig. 1f, generated by introducing the fundamental beam ($\lambda = 1,004\text{ nm}$, 18 kW cm^{-2}). A direct comparison of the SHG signal collected from single KNbO₃ and ZnO nanowires is complicated by the anisotropic scattering related to their respective rectangular and hexagonal cross-sections. However, a rough estimate for the d_{eff} of KNbO₃ nanowires based on the relative ratio of integrated signals is possible and we found it to be $\sim 9.1\text{ pm V}^{-1}$. This illustrates that the nonlinear polarizability of KNbO₃ nanowires is larger than that for ZnO, as expected from the consideration of bulk values.

The second key requirement for a versatile nonlinear circuit element for use in nano-photonics is wave mixing, specifically sum frequency generation (SFG) ($\omega_3 = \omega_1 + \omega_2$) and difference frequency generation ($\omega_3 = |\omega_1 - \omega_2|$). Fig. 1g shows SFG signals ($\lambda = 423\text{ nm}$, 454 nm) and SHG signals ($\lambda = 525\text{ nm}$, 700 nm), obtained from a single KNbO₃ nanowire by introducing fundamental beams at a variety of different frequencies via the tunable femtosecond pump. This demonstrates the ability of nanowire frequency converters to create four different waves from two fundamental input frequencies ω_1 and ω_2 : $2\omega_1$, $2\omega_2$, and $|\omega_1 \pm \omega_2|$. We did not observe the difference frequency generation signal corresponding to the SFG at 423 nm here because the expected wavelength ($\lambda = 7,200\text{ nm}$) is outside current instrumental limits. The SHG signal at 400 nm was weak, owing to photoabsorption within the nanowire. This set of experiments demonstrates the ability of KNbO₃ nanowires to generate continuously tunable and coherent light throughout the visible spectrum via nonlinear wave mixing. This capability, as well as the nanowire's subwavelength cross-section, enables the development of a novel form of scanning light microscopy.

Recently, laser trapping was used to optically manipulate nanowires in closed aqueous chambers^{8,25}. We hypothesized that a single KNbO₃ nanowire may, when optically trapped, be able to double the frequency of the trapping light and then waveguide this locally generated light to its ends. Single KNbO₃ nanowires were optically trapped using a home-built infrared⁸ optical tweezers instrument (Fig. 2a) with the trap wavelength at $1,064\text{ nm}$, a wavelength popular for optical trapping of wet samples owing to the tolerance of living cells to infrared laser irradiation²⁶. We used an electron-multiplying charge-coupled device (CCD) to search for visible light radiating from trapped nanowires.

¹Department of Chemistry, ²Department of Physics, and ³Biophysics Graduate Group, University of California, ⁴Physical Biosciences Division and ⁵Materials Science Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. ⁶Materials Laboratories, Sony Corporation, 4-16-1 Okata Atsugi-shi, Kanagawa 243-0021, Japan.

*These authors contributed equally to this work.

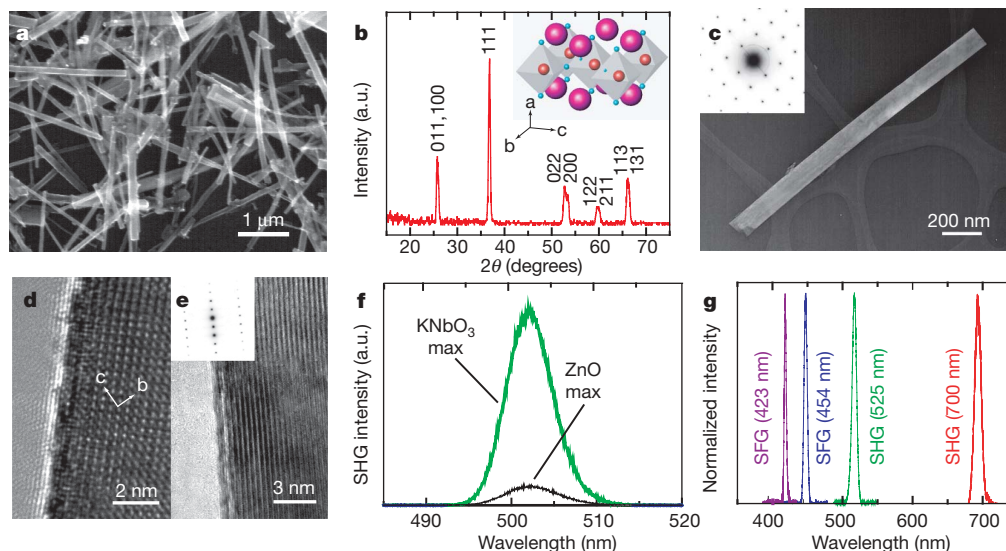


Figure 1 | KNbO₃ nanowires and their structural analysis. **a**, Scanning electron microscopy image of KNbO₃ nanowires. **b**, X-ray powder diffraction pattern of KNbO₃ nanowires. The inset shows the unit cell structure of this material, with spontaneous polarization parallel to the *c*-axis. (a.u., arbitrary units.) **c**, Transmission electron microscope image of a KNbO₃ nanowire and its electron diffraction pattern (inset) with [100] zone

axis and [011] growth direction. **d**, **e**, High-resolution TEM images of single [011] growth-direction KNbO₃ nanowires and an electron diffraction pattern (**e**, inset) with the zone axis of [100] (**d**) and [2–33] (**e**). **f**, Maximum SHG spectra of single KNbO₃ and ZnO nanowires reflecting the larger nonlinear polarizability of KNbO₃. **g**, Panchromatic wavelengths generated by the nonlinear optical processes within individual KNbO₃ nanowires.

As we hoped, light was observed to radiate from the distal end of trapped KNbO₃ nanowires. We charted the radiation profile as a function of position along the nanowire's long axis by coarsely changing the focus of the top objective mounted on a micrometre stage (Fig. 2b). A diffraction-limited spot was observed at the distal end of the wire, revealing optical waveguiding away from the site of photon conversion and emission from the aperture defined by the nanowire's cross-section. Measurements with a colour CCD camera

show that SHG output from the nanowire varies less than 0.2% from the mean over 20 s of collection (Supplementary Fig. 3a).

The light emitted from the nanowire's end was collected through the lower trapping objective and spectrally analysed. Spectral analysis revealed that the light was green, with a wavelength of 531 ± 1.8 nm (Fig. 2c), matching well with the expected SHG signal given a trapping/pump wavelength of 1,064 nm. As a control, we also trapped ZnO and Si nanowires. ZnO (not shown) and Si nanowires (Fig. 2c, black line) did not produce visible light, supporting our conclusion that the green light from the KNbO₃ nanowires was light generated inside the nanowire by SHG and indicating negligible SHG from symmetry breaking at the wire–water interface.

Assuming a non-depleted plane-wave pump geometry (Supplementary Information) and a typical trap irradiance of $\sim 10^8$ W cm⁻², we calculate the total two-photon conversion efficiency $\eta_{2\omega}$ to be at least 10^{-5} . Estimations from this simple theoretical model are in agreement with second harmonic output powers of ~ 10 nW measured with the electron multiplying CCD (Supplementary Information). Although we used a 1,064 nm laser, there is no theoretical limitation to extending both trapping and SHG to the range of wavelengths demonstrated above with femtosecond pulses.

Unlike other nonlinear optical geometries in which alignment is accomplished manually via transducers, here the nonlinear crystal (the nanowire) spontaneously orients itself to the optical axis of the trap/pump laser, resulting in detectable SHG signal along the [011] growth axis. This favourable index matching allows the entire nanowire cavity to participate in the production of second harmonic photons. It is possible that index matching could be further improved by controlling the temperature of the buffer, although we did not explore this here.

Having in hand a nanometric, raster-scannable source of coherent visible light, we wondered whether it could be used to image objects. We used a simple transmission geometry analogous to near-field scanning optical microscopy (NSOM)^{27,28}, in which the sample is scanned through a beam, modulating the fraction of light arriving at a detector. The ultimate resolution of such a transmission microscope depends on the radiation characteristics of the illumination source. In this approach the resolution is of the order of the cross-section of the illumination aperture but many other factors are also important, such as the quality of the probe's end-facets, far-field collection optics and implemented feedback control.

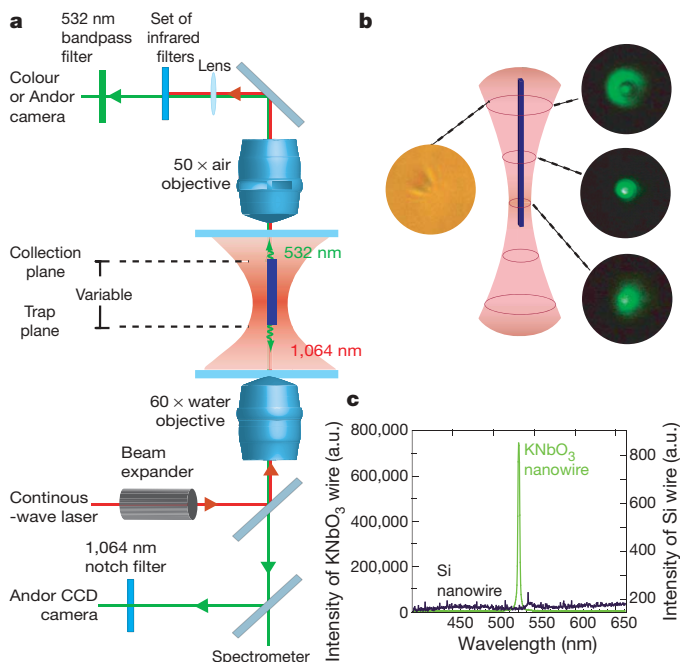


Figure 2 | Radiation from optically trapped single KNbO₃ nanowires. **a**, Detailed set-up for the single-beam optical trapping instrument. **b**, Bright field (left) and SHG (right) images of the trapped KNbO₃ nanowire. Waveguiding of the SHG signal (green) leads to diffraction rings at the distal (top) end of the nanowire which acts as a subwavelength aperture. **c**, Observed spectra for KNbO₃ and Si nanowires. Strong SHG signal at $\lambda = 532$ nm is collected from the trapped KNbO₃ nanowire (green, left axis), while no signal was observed from Si nanowires (black, right axis).

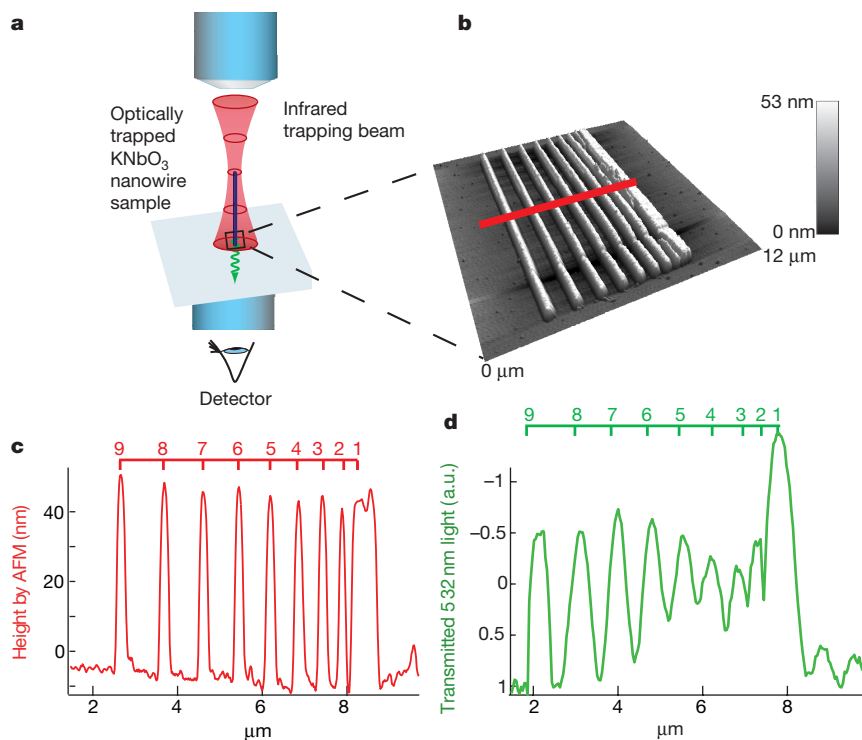


Figure 3 | Transmission line scan of metallic surface pattern with laser trapped KNbO_3 nanowire. **a**, Schematic of inverted optical scanning configuration. **b**, AFM topographic image of thermally evaporated pattern of gold stripes on a glass coverslip. **c**, AFM line scan from region indicated in **b**. **d**, Optical transmission profile captured by scanning a single KNbO_3 nanowire over the metallic surface structure. The nanowire used to create the transmission line scan was measured by AFM to have these dimensions: width = 122 nm, length = 1.4 μm and height = 53 nm. The peak index (there are nine lines in total being measured) is shown at the top of **c** and **d**.

To see whether radiating nanowires can be used to image objects and to characterize the resolution of such a ‘nanowire scanning microscopy’ (Fig. 3a), we fabricated a test pattern via electron beam lithography consisting of a series of 200-nm-wide, 50-nm-thick lines

of gold on transparent glass coverslips with decreasing pitch between each line. A ‘benchmark’ image of the gold pattern was obtained by atomic force microscopy (AFM; Fig. 3b) and shows the progressively finer separation between each line, decreasing from ~1,000 to ~200 nm (Fig. 3c).

The laser trap was then used to raise a single KNbO_3 nanowire to the top coverslip surface with the gold pattern while a piezoelectric stage with nanometre positioning resolution was used to scan the pattern with respect to the wire. Each time the trapped wire passed over a gold feature there was a reduction of transmitted second harmonic emission. Scanning the nanowire tip directly across the surface resulted in a transmitted intensity map with local spatial resolution of the order of the wire’s diameter (Fig. 3d). Measurements of the pitch between lines within the gold pattern were made for both AFM and nanowire optical transmission line scans and agreed to within ~10% (Supplementary Table 1). As expected, the minimum resolvable feature-size (pitch) decreases as the cross-section of the nanowire-probe decreases (Supplementary Fig. 2c, f).

The overall resolution in these experiments is reduced by the use of far-field optics for image collection and (presumably) the fluctuations of the nanowire in the optical potential. Root-mean-square lateral displacements of the wire owing to thermal fluctuations in the optical potential are estimated to be ~10 nm by measurement of the power spectrum of a trapped wire using a position-sensitive photodetector⁸ (Supplementary Fig. 1d); longitudinal fluctuations will be larger owing to reduced confinement of the wire along the optical axis. Indeed, one of the most urgent next steps in the development of nanowire scanning microscopy will be to find ways of reducing longitudinal fluctuations; a theoretically ideal probe would be tapered, because this would increase optical confinement and also the crystal volume in the region of largest field intensity.

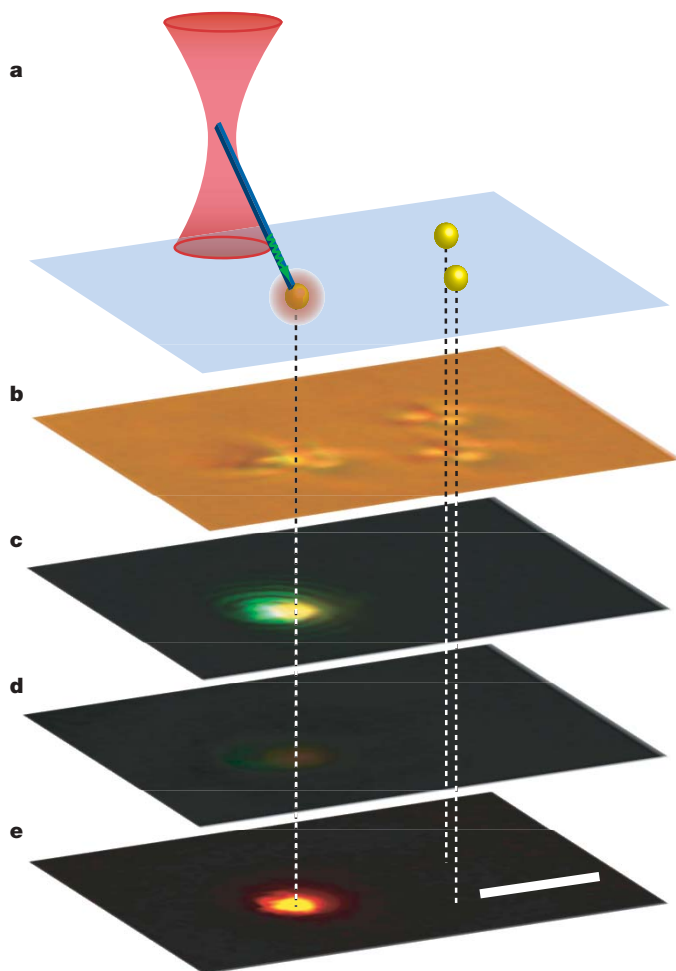


Figure 4 | POPO-3 bead excitation by waveguided SHG signal from an optically trapped KNbO_3 nanowire. **a**, Inverted schematic drawing of the experimental configuration. **b**, Bright-field optical image of the beads, with the nanowire in contact with the leftmost bead. **c**, Colour CCD fluorescence image of **b**, showing green light emission from the nanowire and the orange emission from the bead. **d**, Control image of the same beads with infrared radiation in the plane from **b**, but without a trapped nanowire. **e**, Digital subtraction of **d** from **c**. Scale bar, 3 μm .

In addition to exploring this scanning transmission mode, it is also possible to investigate this new scanning probe in a fluorescence mode. We optically trapped a single KNbO_3 nanowire and touched its distal end to a fluorescent bead (Fig. 4a). In some instances, wires adhered to the bead at the contact point presumably owing to the bead's carboxylate surface functionalization (for adhesion to the top coverslip). Excitation occurred in this near-field geometry that generated a distinct orange fluorescence at the contact point (Fig. 4c). Removal of the nanowire reduced orange emission from the bead by more than 80-fold (Fig. 4d), indicating that the 532 nm SHG emitted by the nanowire was the predominant source of excitation relative to two-photon fluorescence. Figure 4e shows the intensity difference between Fig. 4c and d, obtained by digital subtraction of the red components of the images, which displays the distinctive fluorescence emission due to SHG excitation from the optically trapped KNbO_3 nanowire.

The next step in the development of nanowire-scanning microscopy will be to adapt the now highly refined signal-processing and signal-deconvolution algorithms originally developed for AFM and NSOM to this new method. Although already providing promising lateral resolution (Fig. 3), the utility of nanowire scanning microscopy will further increase once it is understood how to convolve and best interpret the various readouts (fraction of excitation light transmitted, sample fluorescence emission), and the mechanical deflection of the nanowire as it is tapped over a sample, as measured from scattering between the trapping laser^{29,30} and the nanowire. Similar to NSOM, nanowire scanning microscopy is an inherently mechano-optical form of microscopy because the mechanical probe (the nanowire) also functions as a source of coherent visible radiation. One potential advantage of nanowire scanning microscopy is its possible extension to massively parallel excitation arrays using holographic optical elements²⁵ that would be compatible with sealed environments, such as microfluidic chambers.

METHODS SUMMARY

KNbO_3 nanowires were synthesized using a hydrothermal method²². For nonlinear wave mixing measurements, SFG spectra (Fig. 1f, g) were collected by introducing two fundamental beams with the same polarization directions¹¹. Nanowire optical trapping was performed according to ref. 8. The continuous-wave infrared laser ($\lambda = 1064$ nm, ~ 1 W) was introduced from the bottom side of the chamber to trap the nanowire as well as to generate the second harmonic wave at $\lambda = 532$ nm. The trapping instrument was modified with a colour CCD (Fig. 2a) to acquire images of optically trapped KNbO_3 wires at various focal planes by moving the top objective along the optical axis without translation of the trapping point. Spectra were taken through the bottom objective. Gold surface-patterns were made for the scanning measurements using glass coverslips coated with a 2 nm chromium bonding layer. Electron beam lithography was used to define ~ 200 -nm-wide lines with variable pitch. Thermal evaporation of gold yielded lines of gold ~ 50 nm thick. The transmitted SHG signal was collected with a high-speed electron-multiplying CCD camera (Andor, iXon), and custom software written in C++ was used to control the scan size and speed of the xyz-piezo stage. Fluorescent polystyrene beads containing POPO-3 dye molecules were fixed to the surface of a glass coverslip via carboxylate surface functionalization. The distal end of the trapped nanowire was brought in contact with one of the beads (Fig. 4a–c) and colour CCD images were taken (Photometrics, CoolSNAPc).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 20 October 2006; accepted 11 May 2007.

- Sanchez, E. J., Novotny, L. & Xie, X. S. Near-field fluorescence microscopy based on two-photon excitation with metal tips. *Phys. Rev. Lett.* **82**, 4014–4017 (1999).
- Inoué, Y. & Kawata, S. Near-field scanning optical microscope with a metallic probe tip. *Opt. Lett.* **19**, 159–161 (1994).
- Donnert, G. *et al.* Macromolecular-scale resolution in biological fluorescence microscopy. *Proc. Natl Acad. Sci. USA* **103**, 11440–11445 (2006).
- Gustafsson, M. G. L. Nonlinear structured-illumination microscopy: Wide-field fluorescence imaging with theoretically unlimited resolution. *Proc. Natl Acad. Sci. USA* **102**, 13081–13086 (2005).

- Betzig, E. *et al.* Imaging intracellular fluorescent proteins at nanometer resolution. *Science* **313**, 1642–1645 (2006).
- Ghislain, L. P. & Webb, W. W. Scanning-force microscope based on an optical trap. *Opt. Lett.* **18**, 1678–1680 (1993).
- Florin, E. L., Pralle, A., Horber, J. K. H. & Stelzer, E. H. K. Photonic force microscope based on optical tweezers and two-photon excitation for biological applications. *J. Struct. Biol.* **119**, 202–211 (1997).
- Pauzauskie, P. J. *et al.* Optical trapping and integration of semiconductor nanowire assemblies in water. *Nature Mater.* **5**, 97–101 (2006).
- Yang, P. The chemistry and physics of semiconductor nanowires. *MRS Bull.* **30**, 85–91 (2005).
- Sirbul, D. J., Law, M., Yan, H. Q. & Yang, P. D. Semiconductor nanowires for subwavelength photonics integration. *J. Phys. Chem. B* **109**, 15190–15213 (2005).
- Johnson, J. C. *et al.* Near-field imaging of nonlinear optical mixing in single zinc oxide nanowires. *Nano Lett.* **2**, 279–283 (2002).
- Johnson, J. C., Yan, H. Q., Yang, P. D. & Saykally, R. J. Optical cavity effects in ZnO nanowire lasers and waveguides. *J. Phys. Chem. B* **107**, 8816–8828 (2003).
- Pauzauskie, P. J., Sirbul, D. J. & Yang, P. D. Semiconductor nanowire ring resonator laser. *Phys. Rev. Lett.* **96**, 14903 (2006).
- Qian, F. *et al.* Gallium nitride-based nanowire radial heterostructures for nanophotonics. *Nano Lett.* **4**, 1975–1979 (2004).
- Huang, Y., Duan, X., Wei, Q. & Lieber, C. M. Directed assembly of one-dimensional nanostructures into functional networks. *Science* **291**, 630–633 (2001).
- Duan, X. F., Huang, Y., Agarwal, R. & Lieber, C. M. Single-nanowire electrically driven lasers. *Nature* **421**, 241–245 (2003).
- Kind, H., Yan, H. Q., Messer, B., Law, M. & Yang, P. D. Nanowire ultraviolet photodetectors and optical switches. *Adv. Mater.* **14**, 158–160 (2002).
- Law, M. *et al.* Nanoribbon waveguides for subwavelength photonics integration. *Science* **305**, 1269–1273 (2004).
- Shoji, I., Kondo, T., Kitamoto, A., Shirane, M. & Ito, R. Absolute scale of second-order nonlinear-optical coefficients. *J. Opt. Soc. Am. B* **14**, 2268–2294 (1997).
- Zysset, B., Biaggio, I. & Gunter, P. Refractive indexes of orthorhombic KNbO_3 . 1. Dispersion and temperature-dependence. *J. Opt. Soc. Am. B* **9**, 380–386 (1992).
- Kudo, K., Kakiuchi, K., Mizutani, K. & Fukami, T. Characterization of KNbO_3 crystal by traveling solvent floating zone (TSFZ) method. *Jpn. J. Appl. Phys. Part 1* **42**, 6099–6101 (2003).
- Magrez, A. *et al.* Growth of single-crystalline KNbO_3 nanostructures. *J. Phys. Chem. B* **110**, 58–61 (2006).
- Biaggio, I., Kerkoc, P., Wu, L. S., Gunter, P. & Zysset, B. Refractive indexes of orthorhombic KNbO_3 . 2. Phase-matching configurations for nonlinear-optical interactions. *J. Opt. Soc. Am. B* **9**, 507–517 (1992).
- Knutson, K. P., Messer, B. M., Onorato, R. M. & Saykally, R. J. Chirped coherent anti-Stokes Raman scattering for high spectral resolution spectroscopy and chemically selective imaging. *J. Phys. Chem. B* **110**, 5854–5864 (2006).
- Agarwal, R. *et al.* Manipulation and assembly of nanowires with holographic optical traps. *Opt. Express* **13**, 8906–8912 (2005).
- Ashkin, A., Dziedzic, J. M. & Yamane, T. Optical trapping and manipulation of single cells using infrared-laser beams. *Nature* **330**, 769–771 (1987).
- Pohl, D. W., Denk, W. & Lanz, M. Optical stethoscopy—image recording with resolution $\lambda/20$. *Appl. Phys. Lett.* **44**, 651–653 (1984).
- Betzig, E., Trautman, J. K., Harris, T. D., Weiner, J. S. & Kostelak, R. L. Breaking the diffraction barrier - optical microscopy on a nanometric scale. *Science* **251**, 1468–1470 (1991).
- Florin, E.-L., Pralle, A., Stelzer, E. H. K. & Hörber, J. K. H. Photonic force microscope calibration by thermal noise analysis. *Appl. Phys. A* **66**, 75–78 (1998).
- Grier, D. G. A revolution in optical manipulation. *Nature* **424**, 810–816 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported in part by the Dreyfus Foundation and the US Department of Energy (P.Y.), the University of California, Berkeley (J.L.), the Experimental Physical Chemistry Program of the National Science Foundation, and the NASA SRLDA program (R.M.O. and R.J.S.). Y.N. thanks SONY for a research fellowship and P.J.P. thanks the NSF for a graduate research fellowship. Work at the Lawrence Berkeley National Laboratory was supported by the Office of Science, Basic Energy Sciences, Division of Materials Science of the US Department of Energy. We thank T. Kuykendall for transmission electron microscope observations and the National Center for Electron Microscopy for the use of their facilities, L. Sohn for AFM facilities, N. Switz for comments on the manuscript and W. Liang for microfabrication of gold patterns.

Author Contributions Y.N. performed the synthesis and structural characterization of the KNbO_3 wires. Y.N. and R.M.O. designed, performed and analysed the wave mixing experiment. P.J.P. and A.R. designed, performed and analysed the laser trapping and nanoprobe imaging experiments.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.L. (liphardt@physics.berkeley.edu) or P.Y. (p_yang@berkeley.edu).

METHODS

Materials preparation. Potassium hydroxide and niobium pentoxide were mixed with deionized water for 2 h at room temperature²². The slurry was transferred to Teflon vessels and then heated at 150 °C for 2–6 days using the stainless autoclaves. Sizes could be varied by adjusting the reaction time. Scanning electron microscopy images showed that the products are collections of rectangularly shaped nanowires with widths ranging from 40 to 400 nm and the lengths from 1 to 20 µm (Fig. 1a). The products were identified as the single-phase orthorhombic KNbO_3 (*Amm*2, $a = 0.3984$ nm, $b = 0.5676$ nm and $c = 0.5697$ nm) from X-ray powder diffraction measurements (Fig. 1b). Transmission electron microscope images and electron diffraction measurements (Fig. 1c–e) show that the nanowires grow parallel to [011] directions. Twinned structures were observed in some nanowires.

Nonlinear optical measurements. Single nanowire suspensions were obtained after sonication in iso-propanol and placed on a transparent glass coverslip with a nanomanipulator. Femtosecond pulses generated by a regeneratively amplified Ti:sapphire oscillator (wavelength: $\lambda = 800$ nm, 90 fs, 1 kHz) were used to pump an optical parametric amplifier where the wavelength is continuously tunable between 1,150–2,600 nm and can be doubled in frequency by a BaB_2O_4 crystal outside the optical parametric amplifier, providing access to shorter wavelengths before being introduced to the nanowire perpendicular to its growth direction (Fig. 1f, g). The beam spot size was about 1 µm, which is larger than the width of the wires but smaller than their lengths. SFG spectra were measured by introducing two fundamental beams with the same polarization directions. Nanowire output with $\lambda = 423$ nm was produced as a result of SFG: $(423 \text{ nm})^{-1} = (800 \text{ nm})^{-1} + (900 \text{ nm})^{-1}$, and other spectra were obtained as SFG: $(454 \text{ nm})^{-1} = (800 \text{ nm})^{-1} + (1,050 \text{ nm})^{-1}$, SHG: $(525 \text{ nm})^{-1} = 2 \times (1,050 \text{ nm})^{-1}$ and $(700 \text{ nm})^{-1} = 2 \times (1,400 \text{ nm})^{-1}$. From the experimental configuration, coefficients are identified as $d_{\text{eff}} = d_{31}$ for the KNbO_3 nanowire, and $d_{\text{eff}} = d_{33}$ for the ZnO nanowire.

Optical trapping. Nanowires were dispersed into deionized water inside the sample chamber³. The continuous-wave infrared laser ($\lambda = 1,064$ nm, ~1 W) was introduced from the bottom side of the chamber to trap the nanowire as well as to generate the second harmonic wave at $\lambda = 532$ nm. The chamber can be moved coarsely with a manual stage, and finely with a three-axis piezostage. Colour CCD images are taken at various focal planes by moving the top objective along the optical axis without translation of the trapping point, whereas spectra were taken through the bottom objective. On the basis of the experimental configuration and material properties, the nonlinear optical coefficients are considered to be $d_{\text{eff}} = 0$ for both ZnO and Si nanowires. The SHG signal from surface inversion symmetry breaking was too small to detect in this study.

Nanowire scanning. Gold line patterns were fabricated using electron beam lithography with glass coverslips coated with a 2 nm chromium bonding-layer. The electron beam was used to define ~200-nm-wide line patterns with variable pitch. After 50 nm of thermal gold deposition, reactive ion etching was used to remove residual chromium from glass to prevent laser heating and bubble formation during scanning measurements. Transmitted SHG signal was collected (at a frame rate of 20 Hz) by using an Andor iXon camera. Custom software written in C++ was used to control the scan size and speed of xyz-piezo stage (Nano-UHV100; Mad City Lab). During raster scanning experiments the microfabricated gold lines modulated the SHG signal transmitted through the coverslip owing to a change of the local optical transparency. Future work to improve feedback during raster scanning might employ high-speed analysis of the variability in SHG output due to changes in phase matching from opto-mechanical deflections of the nanowire. All scanning data sets were processed using custom software Igor Pro 5 (Wavemetrics). After scanning, wires were fixed to the glass surface using non-specific binding. An AFM (Asylum Research MFP3D combined with Nikon inverted microscope) operating in tapping mode was used to image fixed nanowires used in the scanning measurements as well as electron beam substrates following reactive ion etching. All AFM images were obtained with the MikroMasch probes (model number NSC36/NoAl) having a nominal spring constant 0.6 N m^{-1} .

Bead excitation. Polystyrene beads (1 µm in diameter) containing the fluorescent dye POPO-3 (main emission peak ~570 nm, absorption peak ~532 nm, Molecular Probes, Fig. 4) were fixed to the surface of a glass coverslip via carboxylate surface functionalization. The distal end of the trapped nanowire is brought in contact with one of the beads (Fig. 4a–c) and colour CCD images were taken (Photometrics, CoolSNAPcf). The initial position was set using a computer-controlled piezoelectric stage. The stage could be actuated to remove the nanowire. Finally, the empty laser trap was moved to the original position for corresponding two-photon excitation (Fig. 4d). The subtraction of Fig. 4d from Fig. 4c is performed using the red component of the colour images. Spectra in Supplementary Fig. 4 were obtained (Acton, SpectraPro 300i) using a sample chamber filled with a 20 vol.% aqueous POPO-3 dye solution. The luminescence spectrum of SHG-excitation was obtained by subtraction of pure two-photon luminescence (without trapping a nanowire) from spectra with a trapped KNbO_3 nanowire.

LETTERS

Silicon in the Earth's core

R. Bastian Georg^{1,2}, Alex N. Halliday¹, Edwin A. Schauble³ & Ben C. Reynolds²

Small isotopic differences between the silicate minerals in planets may have developed as a result of processes associated with core formation, or from evaporative losses during accretion as the planets were built up. Basalts from the Earth and the Moon do indeed appear to have iron isotopic compositions that are slightly heavy relative to those from Mars, Vesta and primitive undifferentiated meteorites^{1–4} (chondrites). Explanations for these differences have included evaporation during the ‘giant impact’ that created the Moon (when a Mars-sized body collided with the young Earth). However, lithium⁵ and magnesium⁶, lighter elements with comparable volatility^{7–9}, reveal no such differences, rendering evaporation unlikely as an explanation. Here we show that the silicon isotopic compositions of basaltic rocks from the Earth and the Moon are also distinctly heavy. A likely cause is that silicon is one of the light elements in the Earth's core. We show that both the direction and magnitude of the silicon isotopic effect are in accord with current theory¹⁰ based on the stiffness of bonding in metal and silicate. The similar isotopic composition of the bulk silicate Earth and the Moon is consistent with the recent proposal¹¹ that there was large-scale isotopic equilibration during the giant impact. We conclude that Si was already incorporated as a light element in the Earth's core before the Moon formed.

We have measured the Si isotopic compositions for 44 meteorites and lunar and terrestrial samples (Table 1) using high-resolution multiple collector inductively coupled plasma mass spectrometry (MC-ICPMS)¹². The data are plotted on a conventional three-isotope diagram (Fig. 1) and are consistent with a mass-dependent fractionation line with a slope of $\delta^{29}\text{Si} = 0.5178 \times \delta^{30}\text{Si}$. There are no resolvable nucleosynthetic isotopic anomalies. A similar level of nucleosynthetic homogeneity has been reported for many other elements, such as magnesium⁶, and contrasts with the widespread non-mass-dependent heterogeneity in oxygen¹³.

The range of Si isotopic compositions for meteorites is small (Table 1). The means are $\delta^{30}\text{Si} = -0.58 \pm 0.06\text{‰}$ and $\delta^{29}\text{Si} = -0.30 \pm 0.03\text{‰}$ (± 1 standard deviation, s.d.). It is clear from our data that the Solar System as sampled by chondrites is close to being uniform isotopically; the same applies to basaltic achondrites from Mars and Vesta. The possible exceptions are enstatite chondrites and aubrites, which provide a hint of slightly lighter Si. There is no indication of any significant isotopic effect associated with basaltic partial melting.

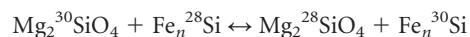
The data for the Earth and the Moon are heavy relative to these meteorite values. The four lunar samples are indistinguishable from each other and give a mean $\delta^{30}\text{Si} = -0.31 \pm 0.03\text{‰}$. The average $\delta^{30}\text{Si}$ of samples from the terrestrial mantle is $-0.38 \pm 0.06\text{‰}$. The few data for clinopyroxene and olivine in mantle xenoliths are similar to the data for basalts, endorsing the view from the meteorite work and the predictions from theory¹⁰ that Si isotope fractionation during partial melting is small. Our data are in excellent agreement with previous estimates for the composition of the bulk silicate Earth (BSE) of $\delta^{30}\text{Si} = -0.4\text{‰}$ based on averages for mafic and ultramafic rocks¹⁴, as well as with previously reported data for the Moon^{15,16} and

meteorites¹⁷ (see Methods). However, the analytical uncertainties in these other studies are too large to resolve differences.

The most obvious distinction between the Earth on the one hand and Mars and Vesta on the other is the size of the body, which would have led to higher average pressures of core formation¹⁸. This is a likely explanation for the heavier silicon isotopic composition of the BSE. It has long been recognized that the Earth's core must contain a significant amount of a light element, candidates for which have included hydrogen, carbon, silicon, sulphur and potassium. A key argument in favour of Si is the high Mg/Si ratio of the silicate Earth^{19,20}. Recent experimental results²¹ are consistent with $\sim 5\text{ wt\%}$ Si in the core¹⁸, which corresponds to a Si deficiency of the BSE of 11%, in accord with its Mg/Si ratio²⁰. The partitioning of Si into liquid iron is thought to be a high-temperature, high-pressure effect²¹ that became important after the Earth had achieved 15% of its current mass¹⁸. Therefore, any resultant metal/silicate Si isotopic fractionation might not be present in smaller objects such as Vesta or Mars.

Some have proposed a ‘heterogeneous accretion’ of the Earth with an early reduced stage²². Silicon would then partition into metal even at low pressures. However, the absence of reduced chondritic meteorites with a chemical composition like that of the Earth has led to these models being largely abandoned. Either way, the isotope data provide evidence that Si is a light element in the core.

Silicon isotope fractionation during core formation is likely because enrichment of heavy isotopes is expected in the phase with the ‘stiffer’ bonds¹⁰, in this case the Si–O bonds in silicates. Sufficient independent constraints exist to provide first order confirmation of the direction and magnitude of such a fractionation. To estimate the likely equilibrium $^{30}\text{Si}/^{28}\text{Si}$ fractionation between silicate and metal, one can examine the reaction:



In this reaction, ^{30}Si substitution is expected to affect free energies primarily by lowering phonon frequencies²³. The size of the isotopic energy shift is calculated by comparing the phonon density of states (PDOS) of each isotopologue. The PDOS of Mg_2SiO_4 is calculated using a first principles model of the forsterite (olivine) structure based on density functional perturbation theory^{24,25}. Olivine is the primary silicate phase in the upper 400 km of the Earth's mantle. The electronic structure is determined using norm-conserving pseudo-potentials and a gradient-corrected functional²⁶. Phonon energies are sampled at one non-zero wave vector, and scaled by 1.043, based on comparison between measured and computed Raman- and infrared-active phonon frequencies (see, for example, ref. 27). Similar models of diopside, quartz and zircon suggest that $^{30}\text{Si}/^{28}\text{Si}$ fractionation between silicates is $<0.3\text{‰}$ above 1,000 °C. Fe_nSi is modelled by analogy with Fe_3Si , which has an atomic packing structure similar to the body-centred cubic phase of iron. The upper limit of the effect of isotope substitution on the phonon energy of Fe_3Si is estimated by assuming that silicon vibrates at the highest frequencies observed in

¹Department of Earth Sciences, University of Oxford, South Parks Road, Oxford OX1 3PR, UK. ²Department of Earth Sciences, ETH Zentrum NW, Clausiusstrasse 25, CH-8092 Zürich, Switzerland. ³Department of Earth and Space Sciences, University of California, Los Angeles, California 90095-1567, USA.

Table 1 | Silicon isotope compositions of samples from the Earth, Moon and meteorites

Sample		$\delta^{30}\text{Si}$	2 s.e.	$\delta^{29}\text{Si}$	2 s.e.	<i>n</i>	SiO_2	Group average			
								(‰ relative to NBS-28)	(wt%)	$\delta^{30}\text{Si}$	1 s.d.
Carbonaceous chondrites											
Allende	CV3	−0.58	0.03	−0.29	0.03	39	35.4	−0.58	0.05	−0.30	0.03
Orgueil	CI	−0.60	0.03	−0.30	0.01	40	22.2				
Murchison	CM2	−0.65	0.03	−0.34	0.03	33	29.7				
ALH 83108	CO3.5	−0.54	0.03	−0.29	0.02	11	34.9				
EET 92002	CK4	−0.53	0.02	−0.27	0.01	11	34.0				
Group average											
Ordinary chondrites											
Cereseto	H5	−0.56	0.03	−0.29	0.02	22	34.4	−0.57	0.04	−0.30	0.03
Cangas de Onis	H5	−0.55	0.03	−0.28	0.03	19	31.6				
Allegan	H5	−0.62	0.03	−0.33	0.03	28	42.1				
Bruderheim	L6	−0.56	0.03	−0.31	0.03	22	43.0				
Bath	H4	−0.51	0.03	−0.26	0.02	20	37.9				
Epinal	H5	−0.60	0.03	−0.32	0.02	22	37.4				
Group average											
Enstatite chondrites											
Abee	EH4	−0.69	0.03	−0.36	0.01	11	46.5				
Eucrites and diogenites											
Sioux Country	Monomict eucrite	−0.52	0.04	−0.25	0.03	20	46.5	−0.56	0.06	−0.29	0.04
Bouvante	Polymict eucrite	−0.53	0.03	−0.26	0.03	22	50.2				
Bereba	Monomict eucrite	−0.50	0.03	−0.26	0.02	22	52.8				
Juvinas	Monomict eucrite	−0.60	0.03	−0.28	0.02	25	50.5				
Pasamonte	Polymict eucrite	−0.66	0.02	−0.35	0.02	22	49.3				
Johnstown UB	Diogenite	−0.56	0.02	−0.32	0.01	11	54.4				
Group average											
Aubrites											
ALH 84009	Aubrite	−0.70	0.03	−0.36	0.02	11	57.4				
Chainpur (LL3.4) chondrules											
XT-1	Chondrule	−0.61	0.04	−0.29	0.01	7	51.4	−0.51	0.19	−0.26	0.11
XT-2	Chondrule	−0.46	0.04	−0.23	0.04	7	38.7				
XT-3	Chondrule	−0.48	0.02	−0.28	0.02	11	36.2				
XT-4	Chondrule	−0.24	0.04	−0.11	0.02	11	32.5				
XT-5	Chondrule	−0.75	0.02	−0.40	0.04	7	41.5				
Group average											
Martian meteorites											
ALH 84001	Orthopyroxenite	−0.57	0.03	−0.30	0.02	12	54.7	−0.59	0.05	−0.31	0.04
Zagami	Basalt	−0.55	0.03	−0.28	0.02	22	47.7				
Nakhla	Olivine clinopyroxenite	−0.64	0.02	−0.35	0.02	13	48.1				
Group average											
Lunar rocks											
15555.115	Olivine-normal basalt	−0.32	0.03	−0.18	0.02	11	35.6	−0.31	0.03	−0.16	0.02
70035.158	High-Ti basalt	−0.30	0.01	−0.16	0.02	11	41.4				
75075.17	High-Ti basalt	−0.35	0.02	−0.17	0.02	11	42.8				
77516.26	High-Ti basalt	−0.27	0.02	−0.14	0.01	11	33.8				
Group average											
Ocean island basalt											
ICE-B1	Scoria Iceland	−0.35	0.04	−0.19	0.02	7	45.4	−0.38	0.06	−0.20	0.04
ICE-B2	Basalt Iceland	−0.35	0.03	−0.21	0.01	9	43.9				
SH-6	Basalt St Helena	−0.31	0.03	−0.14	0.01	11	41.2				
LGG8	Basalt glass Loihi	−0.36	0.03	−0.16	0.01	7	43.9				
P-13	Basalt Principe	−0.49	0.04	−0.26	0.02	11	41.4				
Mid-ocean ridge basalt glasses											
A 127 / D5-5	Atlantic	−0.35	0.04	−0.20	0.01	11	44.5	−0.38	0.06	−0.20	0.04
A 127 / D21-3	Atlantic	−0.51	0.02	−0.26	0.02	11	49.6				
R 93-7	East Pacific Rise	−0.39	0.02	−0.22	0.02	11	48.0				
Cameroon line spinel lherzolites											
C 235A	Olivine	−0.35	0.03	−0.16	0.02	9	NA	−0.38	0.06	−0.20	0.04
C 235 A	Clinopyroxene	−0.41	0.03	−0.23	0.02	11	NA				
C 271 I	Olivine	−0.31	0.03	−0.14	0.03	11	NA				
C 271 I	Clinopyroxene	−0.42	0.04	−0.23	0.03	9	NA				
Group average (last 3 groups)											

All data represent whole rock analyses, except for chondrules and spinel lherzolite mineral separates. The uncertainties on SiO_2 wt% are $\pm 3\%$. Silicon isotopic compositions are reported in parts per thousand of the deviations of the $^{30}\text{Si}/^{28}\text{Si}$ ($J = 30,29$) of the sample relative to the international reference material NBS-28 (ref. 12). Each sample was measured at least 11 times, resulting in an average limiting precision of 0.04‰ and 0.03‰ (2 s.e.) for $\delta^{30}\text{Si}$ and $\delta^{29}\text{Si}$, respectively. The 5 carbonaceous chondrites, 6 ordinary chondrites, 6 HED basaltic achondrites thought to come from Vesta and 3 martian meteorites display identical mean $\delta^{30}\text{Si}$ (± 1 s.d.) of $-0.58 \pm 0.05\text{‰}$, $-0.57 \pm 0.04\text{‰}$, $-0.56 \pm 0.06\text{‰}$ and $-0.59 \pm 0.05\text{‰}$. Even the 5 chondrules, though spanning a considerable range that implicates small scale heterogeneity, yield an identical average of $-0.51 \pm 0.19\text{‰}$. The one enstatite chondrite yields a $\delta^{30}\text{Si}$ of $-0.69 \pm 0.03\text{‰}$ (2 s.e.) at the most negative end of the range found for bulk meteorites. An aubrite (a class of meteorite considered to represent the achondritic derivative of enstatite chondrites) yields an identical value of $-0.70 \pm 0.03\text{‰}$. The four lunar samples are indistinguishable from each other and give a mean $\delta^{30}\text{Si}$ of $-0.31 \pm 0.03\text{‰}$, which is significantly different from the meteorite value. On the basis of the samples analysed here, the terrestrial mantle has an average $\delta^{30}\text{Si}$ of $-0.38 \pm 0.06\text{‰}$, slightly lighter than the lunar data but heavier than the meteorite values. NA, not available.

the structure (10 THz; ref. 28). This gives a rough lower bound of 3.6‰ for the $^{30}\text{Si}/^{28}\text{Si}$ fractionation between silicate and metal at 1,000 °C, and 2.7‰ at 1,200 °C (Fig. 2).

An alternative calculation assumes that silicon oscillates at a frequency corresponding to the root mean square of the total PDOS of Fe_3Si (~7.5 THz), providing a lower limit of the isotope substitution energy. This provides evidence that the maximum fractionation is 4.1‰ at 1,000 °C (3.1‰ at 1,200 °C). In both calculations, silicate is the high- $^{30}\text{Si}/^{28}\text{Si}$ phase. The models considered here give isotope effects (ε in ‰, where $\varepsilon = 1,000 \times \ln \alpha_{\text{silicate-metal}}$ and $\alpha_{\text{silicate-metal}} = (^{30}\text{Si}/^{28}\text{Si})_{\text{silicate}} / (^{30}\text{Si}/^{28}\text{Si})_{\text{metal}}$) of about 1.7 and 1.5 at ~2,000 K (Fig. 2). However, the calculated fractionation factors can only be extrapolated to such higher temperatures if one assumes that the Fe–Si bonding environment does not change greatly on melting.

Pressure could also change the fractionation. All models are calculated at $\sim 10^5$ Pa pressure, under the assumption of harmonic phonon oscillation. Equilibration of metal in the presence of high-pressure silicate phases (for example, MgSiO_3 -perovskite)²¹ might be expected to cause less isotopic fractionation, owing to longer, more ionic Si–O bonds in octahedral coordination sites.

These preliminary estimates can be compared with the fractionation factors calculated from the Si isotope data (Fig. 3). Assuming $\delta^{30}\text{Si}$ for the total Earth of –0.58‰ (average chondrites) and for the BSE of –0.38‰, one can use the BSE Si deficiency of 11% (ref. 20)

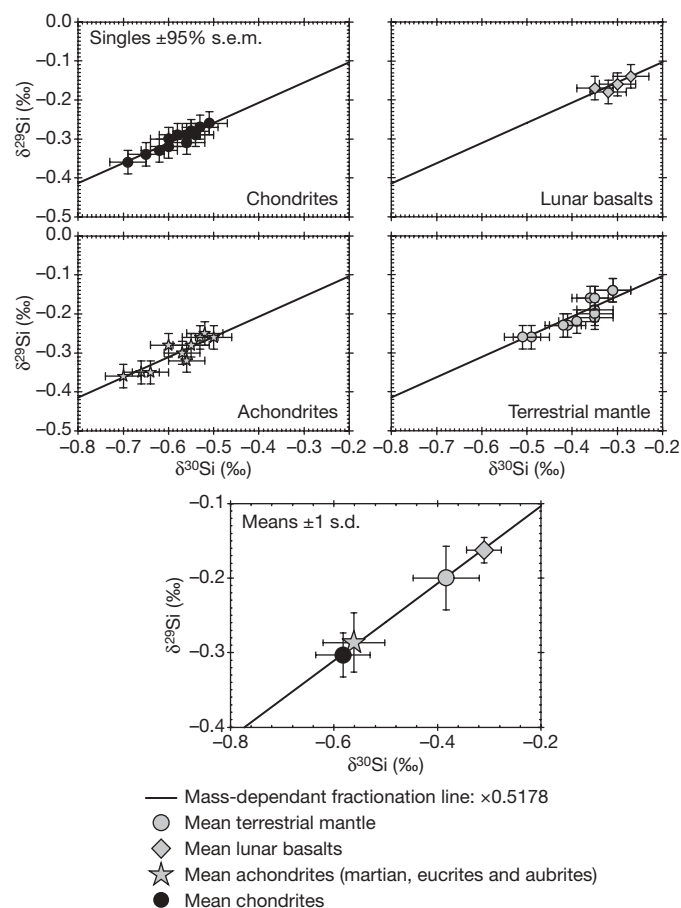


Figure 1 | Silicon isotopic compositions of different Solar System objects. Three-isotope plots showing the data given in Table 1, and divided to illustrate the similarities and differences between reservoirs. It can be seen that all of the data plot on the same mass-dependent fractionation line, consistent with a single homogeneous inner Solar System Si isotopic reservoir. The means and standard deviations for each type are shown in the bottom panel. Chondrites and achondrites are indistinguishable but different from mantle samples from the Earth, which overlap with the data for lunar basalts.

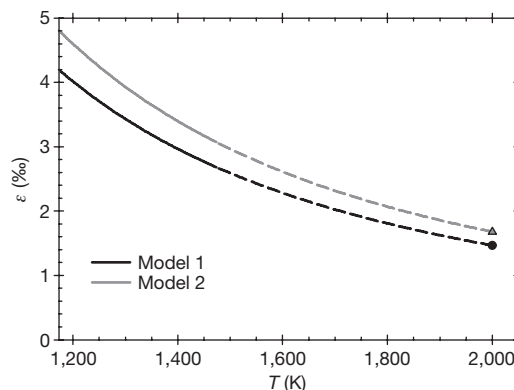


Figure 2 | Theoretical fractionation of silicon isotopes between silicate and metal as a function of temperature. Equilibrium $^{30}\text{Si}/^{28}\text{Si}$ fractionation ($\varepsilon = 1,000 \ln \alpha_{\text{silicate-metal}}$) between silicate and metal is estimated from lattice-dynamics modelling of Mg_2SiO_4 -olivine and body-centred cubic Fe_3Si . The lower (black) curve is calculated assuming Si-atom vibration at 10 THz in the Fe_3Si structure. The upper (grey) curve assumes Si-atom vibration at 7.5 THz in Fe_3Si . The dashed curves at temperatures above 1,200 °C (1,473 K) indicate extrapolation beyond the 1 atm melting point of Fe_3Si .

and a bulk core concentration of 5 wt% Si (ref. 18) to determine the fractionation factor. Both steady-state and Rayleigh-type models can be investigated, and these yield isotope effects (ε) of 1.74‰ and 1.64‰, respectively (Fig. 3), in excellent agreement with our first order calculations (Fig. 2).

There are parallels between the results presented here for Si and those previously reported for Fe. Basalts from the silicate Earth and Moon have Fe that is isotopically heavy by at least 30 p.p.m. per atomic mass unit relative to that of basalts from Mars and Vesta^{1–4}, which is in turn identical to data for chondrites^{1,4}. The pattern is therefore analogous to that found for Si. Whether the mechanisms

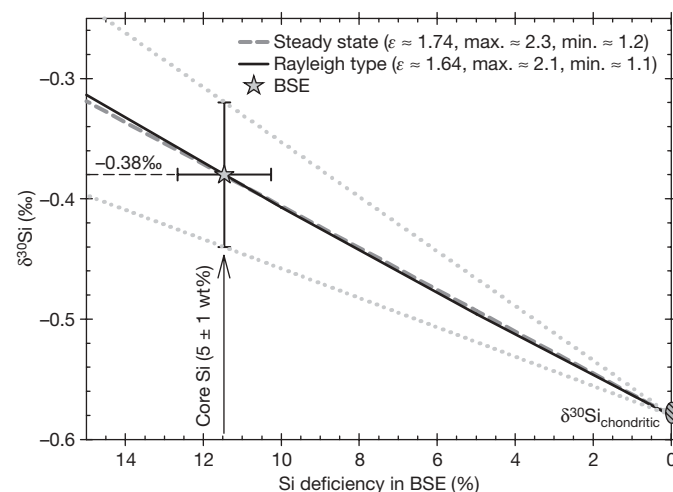


Figure 3 | Fractionation of silicon isotopes by core formation calculated from the composition of the bulk silicate Earth (BSE). Isotope enrichment factors (ε) derived from the measured Si isotope data, as shown for a Rayleigh-type (solid line) or steady-state (dashed line) fractionation scenario. The modelling shows that an enrichment factor of 1.74 (steady state) or 1.65 (Rayleigh type) would suffice to produce the $\delta^{30}\text{Si}_{\text{BSE}}$ of –0.38‰ (star), starting with a $\delta^{30}\text{Si}_{\text{chondritic}}$ of –0.58‰ and taking the Si deficiency of the BSE into account. This would require the core to have ~5% of Si by mass. The isotope effects obtained from the measured data are in good agreement with theoretical predictions for metal–silicate fractionation at temperatures of ~2,000 K. Dotted lines indicate lower and upper limits for ε . The error bars on $\delta^{30}\text{Si}_{\text{BSE}}$ (star) indicate ± 1 wt% Si within the Earth's core and ± 1 s.d. on $\delta^{30}\text{Si}$, respectively.

are related is unclear. It is for example debated whether the effect may not reflect a different style of melting on Earth⁴. However, this would not explain the lunar data. One explanation to consider is that the heavier Fe is also a high-pressure effect but resulting from self-oxidation by perovskite–metal fractionation^{3,18}. Formation of ferric iron can lead to heavier Fe isotopic compositions²⁹ in oxides and silicates, and this might explain the difference relative to Mars and Vesta, which are both of a size at which little or no perovskite would be stable.

A first order calculation of the expected effect can be made by assuming that perovskite–metal fractionation is the same as that of haematite–metal, that the lower-mantle silicate assemblage comprises a 50:50 mixture of periclase and perovskite, and that the fractionation is increased by high-pressure high-spin to low-spin transitions (see Supplementary Discussion). This gives a net $\epsilon^{57/54}\text{Fe}$ (metal/silicate) of -0.04% , which can be compared with a fractionation factor calculated from the effect in the BSE, assuming that all of the silicate and all of the metal formed in isotopic equilibrium, of -0.12% . These are sufficiently close in terms of magnitude and sign to be interesting, but the model is simplistic and further work is needed before firm conclusions can be drawn (see Supplementary Discussion).

As with Fe, the other explanation to consider for the heavy Si in the BSE is partial loss during high-temperature condensation/evaporation. Magnesium and Si have similar condensation temperatures, whereas Li is more volatile⁷. Evaporation experiments show that Si and Mg have similar volatility, with a stronger isotopic fractionation of Mg (ref. 8). Calcium aluminium refractory inclusions (CAIs) are natural evaporation residues and demonstrate the same thing⁹, with Si isotopic fractionations that are only half as extreme as those of Mg, exactly as predicted by experiments⁸. Therefore, available theoretical, experimental and observational evidence suggests that isotope fractionation resulting from partial condensation/evaporation causes similar, if not larger, effects for the lighter element Mg. However, no effects are found for either Li or Mg.

High-pressure core formation explains the Si isotopic composition of the BSE, but leaves the question of how the Moon acquired a similar composition. The Earth and Moon yield oxygen isotopic compositions that are irresolvable at very high precision³⁰, despite widespread Solar System heterogeneity¹³. It was thought that this reflected a similar provenance of material making up the Earth and the planet that struck it in the giant impact, Theia³⁰. The unlikelihood of this has recently been pointed out¹¹, and it has been argued that a more viable explanation is that, following the enormously energetic giant impact, the vapour cloud from which the Moon condensed would have had sufficient time to isotopically equilibrate with the proto-Earth¹¹, leading to identical isotopic compositions. This would also explain the Fe and Si isotopic similarities between the Earth and Moon, and it means that the mechanisms that generated these effects for the Earth need not have been operational on the Moon. In fact, the isotopic similarity of the Moon represents powerful vindication of the new equilibration model¹¹, which in turn provides evidence that Si was already partitioned into the Earth's core before the giant impact. Future work with isotopic measurements on the products of high-pressure experiments should permit elucidation of the conditions of core formation at the time of this last major event in the accretion of the Earth.

METHODS

All samples were dissolved using an alkaline flux and were subsequently transferred into a weakly acidic HCl solution. Silicon was quantitatively separated from the silicate cation matrix by ion exchange chromatography on BioRad columns packed with 1.8 ml DOWEX 50W-X12 (200–400 mesh). A detailed description of the fusion and ion exchange procedures is given elsewhere¹². Silicon isotope data were acquired using the high-resolution Nu Instruments Nu1700 MC-ICPMS (ETH Zürich). Eleven samples were processed twice through the chemical procedure, and were analysed during different sessions more than two months apart. Data accuracy during sample runs was monitored using IRMM-018 and a second batch of NBS-28. The high level of

reproducibility is demonstrated by three different splits of one terrestrial bulk silicate rock sample, which has been processed through the entire procedure, including fusion and column chemistry. The three independent procedural repeats give an average $\delta^{30}\text{Si}$ value of $-0.17 \pm 0.10\%$ (± 2 s.d., $n = 30$) and an average $\delta^{29}\text{Si}$ value of $-0.10 \pm 0.10\%$ (± 2 s.d., $n = 30$). Our meteorite results are fully consistent with previously reported less precise data¹⁷. However, comparisons with still older data are more difficult. Not only are earlier data less precise^{15,16}, they were also measured relative to a different standard, the Cal Tech Rose Quartz Standard (RQS), and not relative to NBS-28 as used here. When RQS was initially calibrated, it was found to be $-0.28 \pm 0.18\%$ on the NBS-28 scale¹⁷. However, our own high-precision measurements of the RQS yield $\delta^{30}\text{Si}$ of $-0.02 \pm 0.10\%$ (± 2 s.d., $n = 17$), rendering older and newer data directly comparable.

Received 11 December 2006; accepted 15 May 2007.

- Poitrasson, F., Halliday, A. N., Lee, D.-C., Levasseur, S. & Teutsch, N. Iron isotope differences between Earth, Moon, Mars and Vesta as possible records of contrasted accretion mechanisms. *Earth Planet. Sci. Lett.* **223**, 253–266 (2004).
- Weyer, S. *et al.* Iron isotope fractionation during planetary differentiation. *Earth Planet. Sci. Lett.* **240**, 251–264 (2005).
- Williams, H. M. *et al.* Iron isotope fractionation in iron meteorites: New insights into metal-sulfide segregation and planetary accretion. *Earth Planet. Sci. Lett.* **250**, 486–500 (2006).
- Schoenberg, R. & von Blanckenburg, F. Modes of planetary-scale Fe isotope fractionation. *Earth Planet. Sci. Lett.* **252**, 342–359 (2006).
- Magna, T., Wiechert, U. & Halliday, A. N. New constraints on the lithium isotope compositions of the Moon and terrestrial planets. *Earth Planet. Sci. Lett.* **243**, 336–353 (2006).
- Wiechert, U. & Halliday, A. N. Non-chondritic magnesium and the origins of the inner terrestrial planets. *Earth Planet. Sci. Lett.* **256**, 360–371 (2007).
- Lodders, K. Solar system abundances and condensation temperatures of the elements. *Astrophys. J.* **591**, 1220–1247 (2003).
- Wang, J., Davis, A. M., Clayton, R. N., Mayeda, T. K. & Hashimoto, A. Chemical and isotopic fractionation during the evaporation of the FeO–MgO–SiO₂–CaO–Al₂O₃–TiO₂ rare earth element melt system. *Geochim. Cosmochim. Acta* **65**, 479–494 (2001).
- Clayton, R. N., Hinton, R. W. & Davis, A. M. Isotopic variations in the rock-forming elements in meteorites. *Phil. Trans. R. Soc. Lond. A* **325**, 483–501 (1988).
- Schauble, E. A. Applying stable isotope fractionation theory to new systems. *Rev. Mineral. Geochem.* **55**, 65–111 (2004).
- Pahlevan, K. & Stevenson, D. J. The oxygen isotope similarity between the Earth and Moon – source region or formation process? *Lunar Planet. Sci.* XXXVI, 2382 (2005).
- Georg, R. B., Reynolds, B. C., Frank, M. & Halliday, A. N. New sample preparation techniques for the precise determination of the Si isotope composition of natural samples using MC-ICP-MS. *Chem. Geol.* **235**, 95–104 (2006).
- Clayton, R. N. Oxygen isotopes in meteorites. *Annu. Rev. Earth Planet. Sci.* **21**, 115–149 (1993).
- Douthitt, C. B. The geochemistry of the stable isotopes of silicon. *Geochim. Cosmochim. Acta* **46**, 1449–1458 (1982).
- Epstein, S. & Taylor, H. P. Jr ¹⁸O/¹⁶O, ³⁰Si/²⁸Si, D/H and ¹³C/¹²C studies on lunar rocks and minerals. *Science* **167**, 533–535 (1970).
- Taylor, H. P. Jr & Epstein, S. Oxygen and silicon isotope ratios of lunar rock 12013. *Earth Planet. Sci. Lett.* **9**, 208–210 (1970).
- Molini-Velsko, C., Mayeda, T. K. & Clayton, R. N. Isotopic composition of silicon in meteorites. *Geochim. Cosmochim. Acta* **50**, 2719–2726 (1986).
- Wade, J. & Wood, B. J. Core formation and the oxidation state of the Earth. *Earth Planet. Sci. Lett.* **236**, 78–95 (2005).
- Allègre, C. J., Poirier, J.-P., Humler, E. & Hofmann, A. W. The chemical composition of the Earth. *Earth Planet. Sci. Lett.* **134**, 515–526 (1995).
- Palme, H. & O'Neill, H. St C. in *The Mantle* (ed. Carlson, R. W.) Vol. 2, *Treatise of Geochemistry* (eds Holland, H. D. & Turekian, K. K.) 1–38 (Elsevier-Pergamon, Oxford, 2003).
- Takafuji, N., Hirose, K., Mitome, M. & Bando, Y. Solubilities of O and Si in liquid iron in equilibrium with (Mg,Fe)SiO₃ perovskite and the light elements in the core. *Geophys. Res. Lett.* **32**, L06313, doi:10.1029/2005GL022773 (2005).
- Newsom, H. E. Accretion and core formation in the Earth: evidence from siderophile elements. In *Origin of the Earth* (eds Newsom, H. E. & Jones, J. H.) 273–288 (Oxford Univ. Press, Oxford, 1990).
- Kieffer, S. W. Thermodynamics and lattice vibrations of minerals. 5. Application to phase equilibria, isotopic fractionation and high-pressure thermodynamic properties. *Rev. Geophys. Space Phys.* **20**, 827–849 (1982).
- Baroni, S., Gironcoli, S., Dal Corso, A. & Giannozzi, P. Phonons and related crystal properties from density-functional perturbation theory. *Rev. Mod. Phys.* **73**, 515–562 (2001).

25. Gonze, X. *et al.* First-principles computation of material properties: the ABINIT software project. *Comput. Mater. Sci.* **25**, 478–492 (2002).
26. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
27. Schauble, E. A., Ghosh, P. & Eiler, J. M. Preferential formation of ^{13}C - ^{18}O bonds in carbonate minerals, estimated using first-principles lattice dynamics. *Geochim. Cosmochim. Acta* **70**, 2510–2529 (2006).
28. Randl, O. G. *et al.* Lattice dynamics and related diffusion properties of intermetallics: I. Fe_3Si . *J. Phys. Condens. Matter* **7**, 5983–5999 (1995).
29. Williams, H. M. *et al.* Systematic iron isotope variations in mantle rocks and minerals: the effects of partial melting and oxygen fugacity. *Earth Planet. Sci. Lett.* **235**, 435–452 (2005).
30. Wiechert, U. *et al.* Oxygen isotopes and the Moon-forming giant impact. *Science* **294**, 345–348 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to C. Smith and colleagues at the Natural History Museum, London, for aliquots of most of the meteorite samples and to

CAPTEM for the lunar samples. MORB and Loihi basalt glasses were provided by M. Garcia, C. Langmuir and W. White. We also thank F. Oberli and colleagues for continuing technical assistance for the isotope facility at ETH Zurich, and SNF, PPARC (now STFC), the NSF and Oxford University for providing financial support. We thank S. Nielsen, H. Williams, D. Stevenson and K. Pahlevan for discussion. The manuscript was improved following comments from T. Elliott.

Author Contributions R.B.G. developed the isotopic methods, produced all of the analytical data and most of the tables and figures, and contributed to the modelling and interpretation. A.N.H. conceived the project, organised sample acquisition, contributed to the interpretation and modelling and wrote most of the text. E.A.S. developed the isotopic fractionation theory and associated calculations and interpretations and wrote important sections of the text. B.C.R. developed the isotopic methods with R.B.G. and made critical standard calibration measurements.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.N.H. (alexh@earth.ox.ac.uk).

Sexually antagonistic genetic variation for fitness in red deer

Katharina Foerster¹, Tim Coulson², Ben C. Sheldon³, Josephine M. Pemberton¹, Tim H. Clutton-Brock⁴ & Loeske E. B. Kruuk¹

Evolutionary theory predicts the depletion of genetic variation in natural populations as a result of the effects of selection, but genetic variation is nevertheless abundant for many traits that are under directional or stabilizing selection¹. Evolutionary geneticists commonly try to explain this paradox with mechanisms that lead to a balance between mutation and selection². However, theoretical predictions of equilibrium genetic variance under mutation–selection balance are usually lower than the observed values, and the reason for this is unknown³. The potential role of sexually antagonistic selection in maintaining genetic variation has received little attention in this debate, surprisingly given its potential ubiquity in dioecious organisms. At fitness-related loci, a given genotype may be selected in opposite directions in the two sexes. Such sexually antagonistic selection will reduce the otherwise-expected positive genetic correlation between male and female fitness⁴. Both theory^{5–7} and experimental data^{8–12} suggest that males and females of the same species may have divergent genetic optima, but supporting data from wild populations are still scarce^{13–15}. Here we present evidence for sexually antagonistic fitness variation in a natural population, using data from a long-term study of red deer (*Cervus elaphus*). We show that male red deer with relatively high fitness fathered, on average, daughters with relatively low fitness. This was due to a negative genetic correlation between estimates of fitness in males and females. In particular, we show that selection favours males that carry low breeding values for female fitness. Our results demonstrate that sexually antagonistic selection can lead to a trade-off between the optimal genotypes for males and females; this mechanism will have profound effects on the operation of selection and the maintenance of genetic variation in natural populations.

Polygynous ungulates such as red deer are particularly likely candidates for taxa in which sexually antagonistic genetic variation may occur, because their pronounced size dimorphism and male weaponry require substantial genotype–sex interactions to enable different developmental schemata for males and females. Furthermore, reproductive roles differ greatly between the sexes. Males compete intensely for matings during the short annual rut but do not invest in offspring care, whereas female maternal investment extends over a long period during each reproductive event. Consequently, male and female life histories are likely to be under divergent selective pressures, and a particular genotype may have very different effects on fitness in males than in females. Last, the polygynous mating system, with short periods of tenure by dominant males, probably reduces the correlation for fitness between mating pairs, enhancing the chance that sexually antagonistic genetic variation may evolve¹⁶. We investigated the heritable genetic basis of variation in fitness in

a wild population of red deer living in the North Block of the Isle of Rum, Scotland. We used individual contributions to population growth, estimated through a method known as ‘de-lifing’¹⁷, as a measure of fitness. This measure, $p_{t(i)}$, estimates an individual’s annual contribution to changes in population size through both reproduction and survival, and it approximates the expected future representation of an individual’s alleles in the population gene pool¹⁷. The de-lifing method measures performance at each potential reproductive event, rather than on the basis of a per-generation time scale. This allows the incorporation of additional information about annual environmental variation, as well as data from incomplete life histories. Where complete life histories were available, we also calculated two lifetime measures of fitness, the lifetime sum of $p_{t(i)}$, and lifetime reproductive success (LRS) as the number of surviving offspring produced over the entire lifespan. Methods and results for the traditional fitness measure LRS are given in Supplementary Information. The correlation coefficient between LRS and the lifetime sum of $p_{t(i)}$ was 0.79 and 0.77 in males and females, respectively ($n = 284$ males and 301 females, both $P < 0.001$).

First, we conducted classical parent–offspring regressions to investigate the sex-specific heritability of fitness. We regressed average son and average daughter values of the lifetime sum of $p_{t(i)}$ (only for individuals with complete life history information) on the values for their fathers and mothers (Fig. 1). Male red deer with a relatively high lifetime sum of $p_{t(i)}$ sired, on average, daughters with a relatively low lifetime sum of $p_{t(i)}$ (Fig. 1; $P = 0.003$). However, there was no significant relationship between the lifetime sum of $p_{t(i)}$ of fathers and sons, or of mothers and their offspring (Fig. 1). Although the slope of the mother–daughter regression was not significantly positive, it differed from the slope of the father–daughter regression (test for the equality of regression slopes¹⁸, $P < 0.01$). This may indicate an underlying genetic antagonistic effect, which is masked by compensating maternal effects on daughters of females with low breeding values for fitness. The slopes of the mother–daughter and the mother–son regressions were not significantly different ($P > 0.05$). If female red deer can compensate for their offspring’s disadvantages from sexually antagonistic genes through direct maternal effects, we would indeed not expect such a difference in a phenotypic analysis. Under the assumption that the patterns observed in the parent–offspring regressions are partly due to heritable genetic variance for fitness, we predicted that $p_{t(i)}$ would have a heritable component and that it would show a negative genetic covariance between the sexes.

We estimated heritable genetic variance by using the ‘animal model’, a mixed-model approach that uses the relatedness between all pairs of individuals in a pedigree to estimate the genetic variance

¹Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3JT, UK. ²Division of Biology and Centre for Population Biology, Imperial College, Silwood Park, Ascot, Berkshire SL5 7PY, UK. ³Edward Grey Institute, Department of Zoology, University of Oxford, Oxford OX1 3PS, UK. ⁴Department of Zoology, University of Cambridge, Cambridge CB2 3EJ, UK.

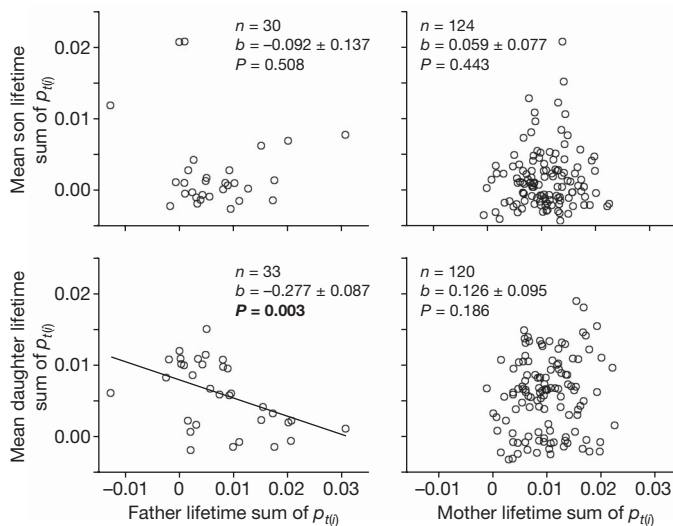


Figure 1 | Sex-specific parent-offspring regressions of observed fitness in red deer. Fitness was estimated as the lifetime sum of p_{ti} , a measure of an individual's annual contribution to changes in population size through reproduction and survival. Here, only deer with complete life history information are included. The very different sample sizes for fathers and mothers are expected under this highly polygynous mating system. The strong skew in male reproductive success produces a female-biased sex ratio among reproductively active red deer. Many males never reproduce and appear as sons, although not as fathers, in this analysis. b is the slope of a linear regression of mean offspring fitness on parental fitness.

component of a trait¹⁹. In contrast with the parent-offspring regressions, this approach allowed us to use all available data in one analysis, to account for direct maternal effects, and to include repeated annual measures of p_{ti} , data from males and females that never reproduced, and data from incomplete life histories. As expected for a fitness-related trait^{20,21}, the annual measure of p_{ti} showed low narrow-sense heritability h^2 (significant only in females; Table 1), but high coefficients of additive genetic variation. The lower heritability in males may be due to high stochasticity of male mating success (increasing the residual variance) and the lack of information on some paternities in and around the study area (reducing the genetic variance). However, we detected a significant negative genetic covariance for male and female p_{ti} in a bivariate animal model, which results in a significant negative genetic correlation between male and female p_{ti} (Table 1, $P = 0.002$). Hence, male red deer that contribute most to the annual population growth have female relatives that contribute less, whereas males that are less successful have relatively more successful female relatives. Fitting separate animal models for the two components of p_{ti} , namely annual reproduction and survival, we found a negative genetic correlation between male and female reproduction but not between male and female survival (Table 1). This is consistent with expectation, as sexual antagonism is likely to be more pronounced with respect to reproduction (where

the sexes diverge most) as opposed to survival. We found no significant heritability or genetic covariance for a more crude measure of fitness, LRS (see Supplementary Information). However, the genetic correlation of male and female LRS was also negative and significantly less than +1 (Supplementary Information, $P < 0.001$). When pleiotropic effects are perfectly correlated across the sexes, the expected genetic correlation between male and female traits is of +1 (ref. 4). A significant deviation from this may be caused by antagonistic genetic variation, by sex-biased gene expression, or by both. In either case, it suggests the potential for constraints on the evolution of a single genetic optimum⁴.

We detected a significant positive covariance for maternal effects on male and female p_{ti} (Table 1), indicating that a mother's maternal effects on her offspring (that is, those acting in addition to the additive genetic effects of the genes that the offspring has inherited from her) benefit her sons and daughters in similar ways. This positive covariance may possibly have masked any sexually antagonistic genetic effects in our mother-offspring phenotype regressions. In the presence of sexually antagonistic genetic variation for fitness, maternal investment that enhances the fitness of both sons and daughters equally may be particularly adaptive, if mothers are not capable of choosing specific mates or manipulating offspring sex to optimize the genetic merit (or breeding value) for fitness of their sons and daughters. Correlated maternal effects on male and female offspring may therefore be a common reason why sexually antagonistic variation for fitness-related traits is difficult to detect in the field. Laboratory experiments on the fruitfly *Drosophila melanogaster*, however, demonstrated the negative consequences of an antagonistic effect on both male and female phenotypes⁶, suggesting small or uncorrelated maternal effects on offspring fitness in this system.

Our findings show that optimal genotypes differ between male and female red deer, because a genotype that produces a male phenotype with relatively high fitness will, on average, produce a phenotype with lower fitness when expressed in a female. The sexually antagonistic genetic variation for fitness should counteract selection on the optimal male and female genotype, respectively. To test this, we calculated the prediction of each individual's additive genetic merit for the annual measure of p_{ti} in the opposite sex, known as its breeding value⁴, from the animal model. Note that we were able to predict a breeding value for p_{ti} in the opposite sex for all animals in which we observed p_{ti} in opposite-sex relatives. If optimal genotypes differ between the sexes, we expect that males with a high breeding value for female p_{ti} will show low fitness, and vice versa. In support of this expectation, we found that a male's breeding value for female p_{ti} was negatively related to his own phenotypic value of male p_{ti} , and hence fitness (Fig. 2a; $P = 0.017$). This confirms our results from the parent-offspring regression: successful males sired, on average, daughters that were less successful, because their genotype had a lower breeding value for female p_{ti} . In contrast, a female's breeding value for male p_{ti} was not significantly related to her own observed p_{ti} (Fig. 2b). Again, this confirms the finding that successful females did not, on average, produce sons that were less successful. Our study

Table 1 | Variance components, heritability h^2 , and the inter-sexual genetic correlation of fitness in red deer

Variance components No. of observations (individuals)	Mean \pm s.d.	Additive genetic	Maternal	Permanent environment	Residual	h^2	P	CV _A
Female p_{ti} 2852 (387)	0.00080 \pm 0.0013	12.53 (3.46)	0.30 (2.11)	0	132.74 (3.76)	0.086 (0.023)	0.002	44.25
Male p_{ti} 2116 (342)	0.00018 \pm 0.0016	7.93 (7.06)	1.03 (4.24)	12.76 (7.17)	205.02 (6.82)	0.035 (0.031)	0.128	156.45
Covariances		Additive genetic	Maternal			Genetic correlation	P	
Female p_{ti} – male p_{ti}		-10.26 (4.18)	5.08 (2.01)			-0.95 (0.42)	0.014	
Female reproduction R – male reproduction R		-8.42 (1.97)	3.61 (0.98)			-1.38 (0.42)	<0.001	
Female survival S – male survival S		-1.35 (2.59)	2.02 (1.51)			-0.45 (0.89)	0.527	

Univariate animal models were used to partition the variance of male and female p_{ti} , a measure of an individual's annual contribution to changes in population size through reproduction (R) and survival (S). Covariances and correlations of female and male p_{ti} and its components R and S were obtained from bivariate animal models. Values in parentheses are standard errors as estimated by the software ASReml. Statistically significant components ($P < 0.05$) are indicated in bold; the probabilities shown are for additive genetic effects. All presented variance components and covariances are based on transformed values of p_{ti} (R and S ($\times 10,000$)). The coefficient of variation for the additive genetic component (CV_A) is a standardized measure of genetic variance and was calculated as $100 \times \sqrt{(\text{additive genetic variance}) / (\text{transformed mean } p_{ti})}$.

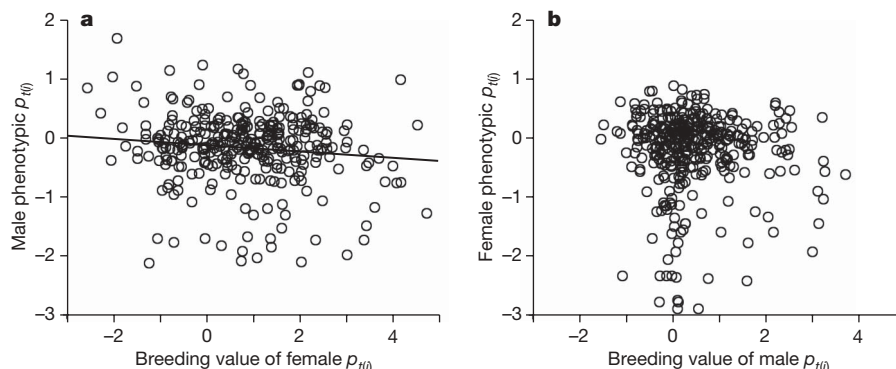


Figure 2 | Selection on opposite-sex breeding values of annual fitness in males and females. The plotted data and regression line are based on average phenotypic values of residual annual $p_{ti(i)}$ (accounting for effects of age and, in females, recent reproductive history); the statistics refer to linear mixed models with male (a) or female (b) identity as a random factor. a, Relationship

between a male's mean phenotypic value of $p_{ti(i)}$ and his breeding value for female $p_{ti(j)}$ ($n = 332$ males, $b = (-8.62 \pm 3.62) \times 10^{-5}$, Wald statistic = 5.67, d.f. = 1, $P = 0.017$). b, Relationship between a female's mean phenotypic value of $p_{ti(i)}$ and her breeding value for male $p_{ti(j)}$ ($n = 371$ females, $b = (-5.41 \pm 3.60) \times 10^{-5}$, Wald statistic = 2.27, d.f. = 1, $P = 0.132$).

therefore provides evidence for sexually antagonistic genetic variation for $p_{ti(i)}$, a measure of annual survival and reproductive success. Our findings suggest that selection for successful males causes a correlated selective pressure against high female breeding values.

In species with male heterogamety, such as mammals, sexually antagonistic genetic variation is expected to accumulate on the X chromosome^{7,11}. X-linked alleles with new antagonistic mutations should spread rapidly within a population's gene pool if they are recessive and beneficial to males, because such alleles are directly accessible to positive selection in all males that inherit them (because males possess only one X chromosome) but are protected from negative selection in all heterozygous females. In concordance with the predictions for an X-linked effect, we observed a negative covariance for fitness between fathers and daughters but no covariance between fathers and sons (Fig. 1). We did not detect the expected covariances between mothers and sons (negative), and mothers and daughters (positive), but these are likely to be masked by direct maternal effects on offspring fitness, which are apparent in our animal model analyses (Table 1). Finally, antagonistic alleles seem to be selected for in males (Fig. 2a), but on average not selected against in females (Fig. 2b). Our results are therefore broadly consistent with the predictions for an antagonistic effect of recessive alleles at one or several X-linked loci, as observed in *Drosophila*¹¹.

Our findings imply a limit to the adaptive evolution of male and female phenotypes in red deer. Sexually antagonistic selection may thus maintain heritable genetic variance in reproductive traits²⁰, but it may also have a role in a reduced response to directional sexual selection, as observed in male red deer²². Sexually antagonistic effects have largely been neglected in models of sexual selection that relate individual mating decisions to the indirect benefit of 'good genes'²³. Our study provides evidence from a natural population that such 'good genes' may be gender-specific and would not provide equal benefits to sons and to daughters^{14,24}. Sexually antagonistic genetic variation could therefore reduce or even reverse indirect benefits from sexual selection. Females that mate with successful males bear the costs of producing daughters with breeding values for low reproductive output. If the antagonistic effect is X-linked, they will also not gain any indirect benefits through their sons. Choosy females should gain fitness advantages through successful grandsons, but this benefit is likely to be reduced through a generation delay and recombination, and may further be attenuated by the likely reduced reproductive output of granddaughters²⁴. The data from red deer on Rum reveal that sexually antagonistic heritable genetic variance is contributing to trait variance in the wild, and that its consequences for sex-specific fitness are not confined to experimental or laboratory populations, which typically show reduced environmental variance for fitness and weak effects of natural selection. As a consequence, the potential of

sexually antagonistic fitness variation needs careful consideration when investigating natural and particularly sexual selection, and more studies focusing on such effects in natural populations are warranted.

METHODS SUMMARY

We used life history data on individual red deer (*Cervus elaphus*) living in the North Block of the Isle of Rum, Scotland, between 1971 and 2005. All animals in the study population are individually recognizable, and their survival and reproductive success have been monitored^{20,22,25–27}. We obtained a population pedigree including 3,559 animals from eight generations, using microsatellite paternity data²⁸ and rut observations²⁰. For all animals that lived for at least three years, we measured fitness using the de-lifing approach¹⁷, a jack-knifing procedure to calculate an individual's contribution to changes in population size, $p_{ti(i)}$, through both reproduction and survival. $p_{ti(i)}$ estimates the relative performance of an individual in each year, and this procedure therefore allowed us to include incomplete life histories. We also present analyses of the two separate components of $p_{ti(i)}$: annual survival S , and annual reproduction R . For a single lifetime measure of fitness (only for individuals with complete life history information, excluding shot deer or deer that are still alive), we used the sum of all annual $p_{ti(i)}$ values, referred to as the lifetime sum of $p_{ti(i)}$. We present P values from two-tailed tests with $\alpha = 0.05$.

We partitioned phenotypic trait variance into additive genetic variance, maternal variance, permanent environment variance, and residual variance by using the animal model as implemented by the software ASReml²⁹. The animal model uses pedigree information to extract the additive genetic component in a mixed-model framework based on restricted maximum-likelihood (REML) estimation¹⁹. We applied univariate models to estimate narrow-sense heritabilities (h^2) and bivariate models for the calculation of inter-sexual covariances and genetic correlations. We used predictions of the additive genetic effect from the bivariate model to investigate selection on opposite-sex breeding values.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 11 December 2006; accepted 11 May 2007.

- Merila, J., Sheldon, B. C. & Kruuk, L. E. B. Explaining stasis: microevolutionary studies in natural populations. *Genetica* **112**, 199–222 (2001).
- Barton, N. H. & Keightley, P. D. Understanding quantitative genetic variation. *Nature Rev. Genet.* **3**, 11–21 (2002).
- Turelli, M. & Barton, N. H. Polygenic variation maintained by balancing selection: Pleiotropy, sex-dependent allelic effects and $G \times E$ interactions. *Genetics* **166**, 1053–1079 (2004).
- Lynch, M. & Walsh, B. *Genetics and Analysis of Quantitative Traits* (Sinauer, Sunderland, Massachusetts, 1998).
- Hedrick, P. W. Antagonistic pleiotropy and genetic polymorphism: a perspective. *Heredity* **82**, 126–133 (1999).
- Rice, W. R. & Chippindale, A. K. Intersexual ontogenetic conflict. *J. Evol. Biol.* **14**, 685–693 (2001).
- Rice, W. R. Sex chromosomes and the evolution of sexual dimorphism. *Evolution Int. J. Org. Evolution* **38**, 735–742 (1984).
- Chippindale, A. K., Gibson, J. R. & Rice, W. R. Negative genetic correlation for adult fitness between sexes reveals ontogenetic conflict in *Drosophila*. *Proc. Natl Acad. Sci. USA* **98**, 1671–1675 (2001).

9. Fedorka, K. M. & Mousseau, T. A. Female mating bias results in conflicting sex-specific offspring fitness. *Nature* **429**, 65–67 (2004).
10. Rice, W. R. Sexually antagonistic genes: experimental evidence. *Science* **256**, 1436–1439 (1992).
11. Gibson, J. R., Chippindale, A. K. & Rice, W. R. The X chromosome is a hot spot for sexually antagonistic fitness variation. *Proc. R. Soc. Lond. B* **269**, 499–505 (2002).
12. Meagher, T. R. The quantitative genetics of sexual dimorphism in *Silene latifolia* (Caryophyllaceae). 1. Genetic variation. *Evolution Int. J. Org. Evolution* **46**, 445–457 (1992).
13. Forsman, A. Opposing fitness consequences of colour pattern in male and female snakes. *J. Evol. Biol.* **8**, 53–70 (1995).
14. Calsbeek, R. & Sinervo, B. Within-clutch variation in offspring sex determined by differences in sire body size: cryptic mate choice in the wild. *J. Evol. Biol.* **17**, 464–470 (2004).
15. Robinson, M. R., Pilkington, J. G., Clutton-Brock, T. H., Pemberton, J. M. & Kruuk, L. E. B. Live fast, die young: Trade-offs between fitness components and sexually antagonistic selection on weaponry in Soay sheep. *Evolution Int. J. Org. Evolution* **60**, 2168–2181 (2006).
16. Rice, W. R. Dangerous liaisons. *Proc. Natl Acad. Sci. USA* **97**, 12953–12955 (2000).
17. Coulson, T. *et al.* Estimating individual contributions to population growth: evolutionary fitness in ecological time. *Proc. R. Soc. Lond. B* **273**, 547–555 (2006).
18. Sokal, R. R. & Rohlf, F. J. *Biometry* 3rd edn (Freeman, New York, 1995).
19. Kruuk, L. E. B. Estimating genetic parameters in natural populations using the 'animal model'. *Phil. Trans. R. Soc. Lond. B* **359**, 873–890 (2004).
20. Kruuk, L. E. B. *et al.* Heritability of fitness in a wild mammal population. *Proc. Natl Acad. Sci. USA* **97**, 698–703 (2000).
21. Houle, D. Comparing evolvability and variability of quantitative traits. *Genetics* **130**, 195–204 (1992).
22. Kruuk, L. E. B. *et al.* Antler size in red deer: heritability and selection but no evolution. *Evolution Int. J. Org. Evolution* **56**, 1683–1695 (2002).
23. Kokko, H. Fisherian and 'good genes' benefits of mate choice: how (not) to distinguish between them. *Ecol. Lett.* **4**, 322–326 (2001).
24. Pischedda, A. & Chippindale, A. K. Intralocus sexual conflict diminishes the benefits of sexual selection. *PLoS Biol.* **4**, e356 (2006).
25. Clutton-Brock, T. H., Guinness, F. E. & Albon, S. D. *Red Deer: Behavior and Ecology of Two Sexes* (Univ. of Chicago Press, Chicago, 1982).
26. Kruuk, L. E. B., Clutton-Brock, T. H., Rose, K. E. & Guinness, F. E. Early determinants of lifetime reproductive success differ between the sexes in red deer. *Proc. R. Soc. Lond. B* **266**, 1655–1661 (1999).
27. Coulson, T., Albon, S., Guinness, F., Pemberton, J. & Clutton-Brock, T. Population substructure, local density, and calf winter survival in red deer (*Cervus elaphus*). *Ecology* **78**, 852–863 (1997).
28. Marshall, T. C. *et al.* Estimating the prevalence of inbreeding from incomplete pedigrees. *Proc. R. Soc. Lond. B* **269**, 1533–1539 (2002).
29. Gilmour, A. R., Gogel, B. J., Cullis, B. R., Welham, S. J. & Thompson, R. *ASReml User Guide 1.0*. (VSN International, Hemel Hempstead, 2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank F. Guinness, A. Donald, S. Morris and many other project field workers; K. Connaghan, K. Byrne, S. Lewis, D. Nussey and J. Slate for genotyping work; and D. Nussey and A. Wilson for comments on earlier drafts of the manuscript. We thank Scottish Natural Heritage for permission to work on the Isle of Rum and their local staff for help and support. This work was funded by the Natural Environment Research Council, the Royal Society, and a Marie Curie European Fellowship.

Author Contributions The idea for this study originated from discussions between L.E.B.K., B.C.S. and K.F. The data used stem from a long-term study run by T.H.B.C., with involvement from L.E.B.K. and J.M.P.; J.M.P. was also responsible for the molecular paternity analysis. K.F. conducted all data analyses and drafted the manuscript. L.K. helped with the quantitative genetics analyses, and T.C. helped in applying the de-living approach. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to K.F. (foerster@orn.mpg.de) or L.E.B.K. (loeske.kruuk@ed.ac.uk).

METHODS

De-lifing. To measure fitness we used the de-lifing approach¹⁷, a jack-knifing procedure to calculate an individual's contribution to changes in population size, $p_{t(i)}$. $p_{t(i)}$ was suggested as an appropriate fitness measure for stochastic environments^{17,30} such as those encountered by red deer. In Supplementary Information we discuss some relevant properties of $p_{t(i)}$ in comparison with lifetime reproductive success, a more traditional fitness estimate. Here we calculated

$$p_{t(i)} = \frac{s_{t(i)} - \bar{s}_t}{N_t - 1} + \frac{f_{t(i)} - \bar{f}_t}{N_t - 1}$$

for individual i in year t , where $s_{t(i)}$ is a binary variable representing whether individual i survives from year t to $t + 1$, and $f_{t(i)}$ is 0.5 times the number of offspring produced by individual i in year t that survive to year $t + 1$. \bar{s}_t and \bar{f}_t are the means of $s_{t(i)}$ and $f_{t(i)}$, and N_t is the adult population size (males and females aged at least 3 years) in year t . Individuals with negative values of $p_{t(i)}$ did worse than the population mean, whereas individuals with positive $p_{t(i)}$ performed above average. $p_{t(i)}$ estimates the relative contribution of an individual to population growth in each year and thus allowed us to include incomplete life histories. We also present results based on analyses of the two separate components of the above equation: an individual's annual survival

$$S = \frac{s_{t(i)} - \bar{s}_t}{N_t - 1}$$

and its annual reproduction

$$R = \frac{f_{t(i)} - \bar{f}_t}{N_t - 1}$$

For a single lifetime measure of fitness (only for individuals with complete life history information, excluding shot deer or deer that are still alive), we used the sum of all annual $p_{t(i)}$ values, referred to as the lifetime sum of $p_{t(i)}$. We present P values from two-tailed tests with $\alpha = 0.05$.

Animal model analyses. We partitioned phenotypic trait variance into additive genetic variance, maternal variance, permanent environment variance, and residual variance by using the animal model as implemented by the software ASReml²⁹. The animal model uses pedigree information to extract the additive genetic component in a mixed-model framework based on restricted maximum-likelihood (REML) estimation¹⁹. The pedigree of the red deer on Rum included 3,559 animals from eight generations. We modelled $p_{t(i)}$, R and S as repeated annual measures and chose fixed effects that have been shown to influence reproductive success in this population. For both sexes we included the year of birth to account for cohort effects²⁶, and the individual's age as a quadratic function²². For females only, we added population subdivision (to account for variation in habitat quality)²⁷ and recent reproductive history (whether or not the female had reared a calf the previous year and whether or not that calf survived for six months)³¹.

We ran univariate animal models separately for males and females, to calculate narrow-sense heritability (h^2) and to obtain starting values for the bivariate analysis. We estimated the significance of h^2 as the probability that the additive genetic variance component was greater than zero by using a log-likelihood ratio test. We then applied bivariate animal models with the male and the female trait as two dependent variables and estimated the significance of genetic and maternal covariances by comparing the residual deviance of the final model with that of a model with a fixed covariance of 0 in a log-likelihood ratio test. We multiplied $p_{t(i)}$, R and S by 10,000 to enable model convergence, and all reported model parameters are based on transformed values.

Breeding values. An individual's breeding value for a given phenotypic trait is the total additive effect of its genes on that trait⁴. In an animal model, the breeding value is estimated as an individual's best linear unbiased predictor (BLUP) for the additive genetic effect. It is determined by the deviation of the individual's own phenotype and those of all its relatives, scaled by their relatedness to the given individual, from the population mean¹⁹. For all animals in which we observed $p_{t(i)}$ in opposite-sex relatives, we were able to predict a breeding value for $p_{t(i)}$ in the opposite sex. To investigate selection on opposite-sex breeding values of $p_{t(i)}$, we reran the univariate analyses as linear mixed models without the additive genetic effect and included an individual's breeding value for $p_{t(i)}$ in the opposite sex as a fixed effect.

30. Metcalf, C. J. E. & Pava, S. Why evolutionary biologists should be demographers. *Trends Ecol. Evol.* **22**, 205–212 (2007).

31. Clutton-Brock, T. H., Guinness, F. E. & Albon, S. D. The costs of reproduction to red deer hinds. *J. Anim. Ecol.* **52**, 367–383 (1983).

Lateral habenula as a source of negative reward signals in dopamine neurons

Masayuki Matsumoto¹ & Okihide Hikosaka¹

Midbrain dopamine neurons are key components of the brain's reward system¹, which is thought to guide reward-seeking behaviours^{2–4}. Although recent studies have shown how dopamine neurons respond to rewards and sensory stimuli predicting reward^{1,5,6}, it is unclear which parts of the brain provide dopamine neurons with signals necessary for these actions. Here we show that the primate lateral habenula, part of the structure called the epithalamus, is a major candidate for a source of negative reward-related signals in dopamine neurons. We recorded the activity of habenula neurons and dopamine neurons while rhesus monkeys were performing a visually guided saccade task with positionally biased reward outcomes⁷. Many habenula neurons were excited by a no-reward-predicting target and inhibited by a reward-predicting target. In contrast, dopamine neurons were excited and inhibited by reward-predicting and no-reward-predicting targets, respectively. Each time the rewarded and unrewarded positions were reversed, both habenula and dopamine neurons reversed their responses as the bias in saccade latency reversed. In unrewarded trials, the excitation of habenula neurons started earlier than the inhibition of dopamine neurons. Furthermore, weak electrical stimulation of the lateral habenula elicited strong inhibitions in dopamine neurons. These results suggest that the inhibitory input from the lateral habenula plays an important role in determining the reward-related activity of dopamine neurons.

Dopamine neurons in the substantia nigra pars compacta respond to rewards or sensory stimuli that reliably predict the rewards. The response is positive (an increase in activity) or negative (a decrease in activity) if the value of the reward is higher or lower, respectively, than predicted^{1,5,6,8}. However, it is unclear which brain areas provide dopamine neurons with reward-related information. Among many brain areas that project to the substantia nigra pars compacta¹, we decided to investigate the lateral habenula, which is known to project to the substantia nigra pars compacta⁹ (Supplementary Fig. 2) with inhibitory effects on dopamine neurons¹⁰. The lateral habenula has been implicated in anxiety¹¹, stress^{12,13}, pain¹⁴, avoidance learning^{15,16}, attention¹⁷, human reward processing^{18,19}, and psychosis^{20,21}.

To examine the role of the lateral habenula in reward processing, we compared the activity of habenula neurons and dopamine neurons in two monkeys (L and E) performing a visually guided saccade task with positionally biased reward outcomes (hereafter called 'reward-biased visual saccade task')⁷ (Fig. 1a). The target was presented randomly on the right or left and the monkeys had to make a saccade to it immediately. Correct saccades were signalled by tone stimuli 200 ms after the saccades. Saccades to one position were rewarded, whereas saccades to the other position were not rewarded. Thus, the target instructed the saccade direction and indicated the reward contingency (reward or no-reward). In rewarded trials, a liquid reward was delivered which started simultaneously with the tone stimulus. The position-reward contingency was fixed for 24

consecutive trials and was then reversed abruptly for the next block with no external instruction. In this task, the saccade latencies were reliably related to reward contingency. Figure 1b shows the distribution of the saccade latencies of monkey L. Both monkeys showed significantly shorter saccade latencies in rewarded trials than in unrewarded trials (Supplementary Note A).

Single cell activity of 49 lateral habenula neurons (37 in monkey L and 12 in monkey E) was recorded. These neurons were estimated to be in the lateral habenula using MRI and their localization was confirmed histologically (see Supplementary Fig. 3 and Supplementary Note B). Figure 2a shows the activity of a single neuron recorded in the left habenula while the monkey was performing the reward-biased visual saccade task. The activity increased phasically after the appearance of the saccade target indicating the absence of upcoming reward, and decreased after the appearance of the target indicating the presence of upcoming reward. The increase and decrease depended on the reward contingency, regardless of target position.

Many of the 49 lateral habenula neurons behaved similarly to the sample neuron shown in Fig. 2a. In order to evaluate the effect of reward contingency and target position on the response to the saccade target (hereafter called post-target response), we performed

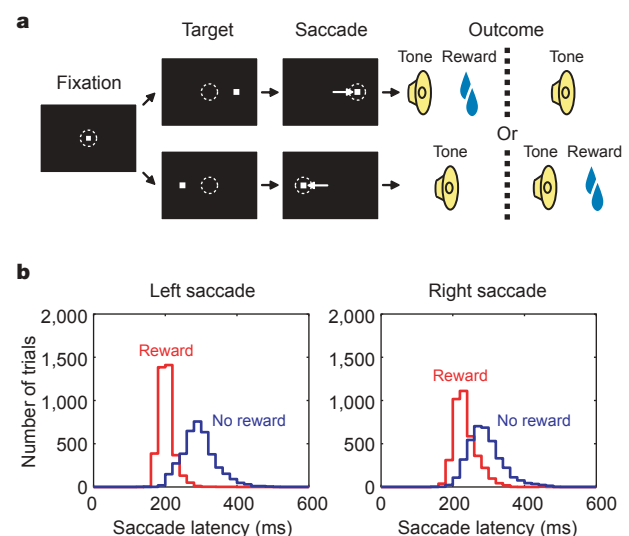


Figure 1 | Behavioural task and monkey's performance. **a**, Sequence of events in the one-direction-rewarded version of the visually guided saccade task (reward-biased visual saccade task). The position-reward contingency was fixed in a block of 24 trials and was reversed in the next block. See text for details of tone and reward. **b**, Distribution of saccade latencies in rewarded trials (red) and in unrewarded trials (blue) (data from monkey L). Saccades in the first trials after the changes in position-reward contingency have been excluded.

¹Laboratory of Sensorimotor Research, National Eye Institute, National Institutes of Health, Bethesda, Maryland 20892-4435, USA.

two-way analysis of variance (ANOVA) for each neuron. Of the 49 neurons, 43 showed a significant main effect of reward contingency, and 10 neurons showed a significant main effect of target position ($P < 0.01$, two-way ANOVA). As shown in Fig. 2b, the reward index (see Methods) was predominantly larger than the position index ($P < 10^{-8}$, Wilcoxon signed-rank test). Even for the 10 neurons that were affected by target position (green dots in Fig. 2b), the position index was smaller than the reward index. Thus, the post-target

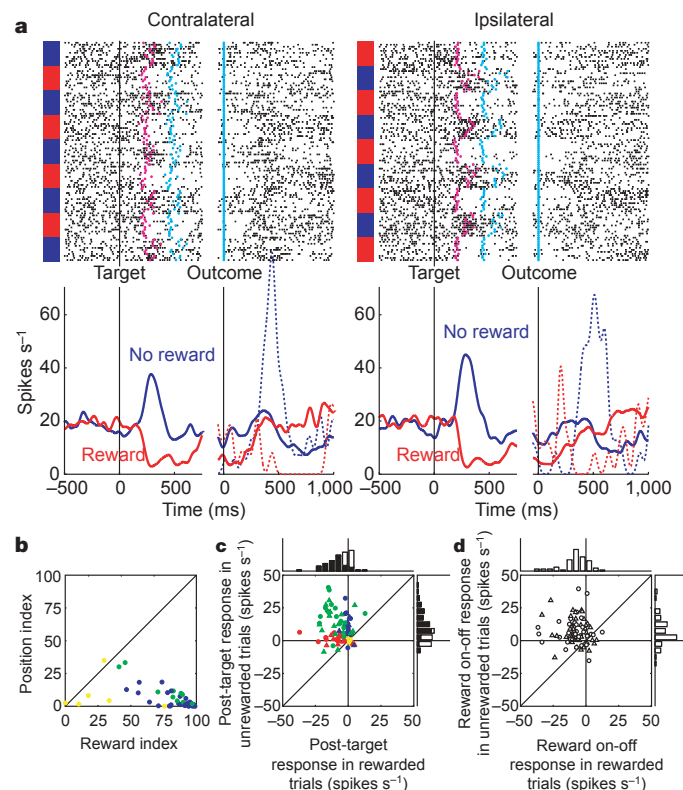


Figure 2 | Activity of habenula neurons during the reward-biased visual saccade task. **a**, Activity of a representative neuron recorded in the left lateral habenula shown separately for contralateral saccades (left) and ipsilateral saccades (right). For each saccade direction, rasters of spikes (dots, top) and spike density functions (SDFs; graphs, bottom) are aligned at the onset of target (left) and at the onset of outcome (tone) (right). The rasters are shown in order of occurrence of trials for each direction from top to bottom. Colour codes to the left of the rasters indicate reward contingency (red, rewarded trials; blue, unrewarded trials). Magenta dots indicate saccade onsets. Cyan dots indicate outcome onsets, which were aligned in the right half of rasters. The SDFs are shown for each reward contingency (red, rewarded trials; blue, unrewarded trials). Continuous curves indicate activity in trials excluding the first trials after the changes in position-reward contingency. Dotted curves indicate activity in the first trials (shown only for outcome). **b**, Comparison between position index and reward index in habenula neurons ($n = 49$). Blue dots indicate neurons with statistically significant main effect of reward contingency ($P < 0.01$, two-way ANOVA). Green dots indicate neurons with statistically significant main effect of both reward contingency and target position ($P < 0.01$). Yellow dots, no significance. **c**, Comparison of post-target responses between unrewarded trials and rewarded trials in habenula neurons ($n = 43$). Blue, red and green dots indicate neurons with statistically significant post-target response in unrewarded trials, in rewarded trials and in both of them, respectively ($P < 0.01$, Wilcoxon signed-rank test). Yellow dots, no significance. Circles and triangles indicate post-target responses to ipsilateral and contralateral target, respectively. The marginal histograms show the distribution of post-target responses for each reward contingency. Black bars indicate neurons with statistically significant post-target responses ($P < 0.01$, Wilcoxon signed-rank test). White bars, no significance. **d**, Comparison of reward on-off responses in first trials after the changes in position-reward contingency between unrewarded trials and rewarded trials in habenula neurons ($n = 43$). Conventions are the same as **c** except that the statistical analysis was not conducted.

response was mainly influenced by the reward contingency rather than target position. The post-target responses of the 43 neurons were predominantly positive (increase in activity) in unrewarded trials and negative (decrease in activity) in rewarded trials (Table 1). This is also evident in the scatter plot in Fig. 2c, which shows that the median post-target response was significantly larger than zero in unrewarded trials and smaller than zero in rewarded trials ($P < 0.01$, Wilcoxon signed-rank test). Control experiments showed that the post-target response was a visual response, not a saccadic response (see Supplementary Fig. 4 and Supplementary Note C).

These properties of habenula neuronal activity were similar to those of dopamine neurons, but in the opposite manner (Fig. 3). We recorded the activity of 62 dopamine neurons in the same monkeys (42 in monkey L and 20 in monkey E). Their activity increased and decreased phasically in response to the saccade target in rewarded and unrewarded trials, respectively, regardless of the target direction (sample neuron activity in Supplementary Fig. 5a, and the average activity in Fig. 3b and Supplementary Fig. 6b). The responses were visual, not saccade-related, confirming a previous report⁶.

Both habenula and dopamine neurons changed their activity similarly after the position-reward contingency was reversed (Fig. 3c and d). After a reward-to-no-reward transition (blue curves), the post-target response in habenula neurons (Fig. 3c, top) increased rapidly (from negative to positive values) whereas the post-target response in dopamine neurons (Fig. 3d, top) decreased rapidly (from positive to negative values). After a no-reward-to-reward transition (red curves), the post-target response in habenula neurons decreased rapidly whereas the post-target response in dopamine neurons increased rapidly. The post-target responses in both habenula and dopamine neurons were then stable for the remainder of the block. The development of the post-target responses in habenula and dopamine neurons (Fig. 3c and d, top) was paralleled by the changes in saccade latency (Fig. 3c and d, bottom), although the changes in the neuronal responses were quicker than the changes in saccade latency especially after a reward-to-no-reward transition.

Habenula neurons also responded differentially to the delivery and omission of reward, as shown in the right half of the raster/spike density functions (SDFs) in Figs 2a and 3a. The responses (hereafter called reward on-off responses) were particularly strong in the first trials after the reversal of the position-reward contingency (dotted lines in the SDFs of Figs 2a and 3a), and appeared as a phasic increase and decrease, respectively, after the omission and delivery of reward. This is evident in the scatter plot in Fig. 2d, which shows that the median reward on-off response was significantly larger than zero in unrewarded trials and smaller than zero in rewarded trials ($P < 0.01$, Wilcoxon signed-rank test). The reward on-off response then declined rapidly (Fig. 3c, middle).

Table 1 | Post-target response of habenula neurons

		Reward trials			
		Excitation	Inhibition	No response	All
Ipsilateral					
No reward trials	Excitation	1(1)	14(11)	12(10)	27(22)
	Inhibition	0(0)	1(1)	1(1)	2(2)
	No response	0(0)	10(4)	4(3)	14(7)
	All	1(1)	25(16)	17(14)	43(31)
Contralateral					
No reward trials	Excitation	3(3)	12(7)	11(10)	26(20)
	Inhibition	0(0)	2(2)	1(1)	3(3)
	No response	2(1)	8(5)	4(2)	14(8)
	All	5(4)	22(14)	16(13)	43(31)

Shown are the numbers of lateral habenula neurons exhibiting a significant excitation, significant inhibition, or no significant response ($P < 0.01$, Wilcoxon signed-rank test). Forty-three neurons that showed a significant main effect of reward contingency in two-way ANOVA comprised the sample used for this analysis. Numbers in parentheses indicate the number of habenula neurons recorded from monkey L; the rest was recorded from monkey E.

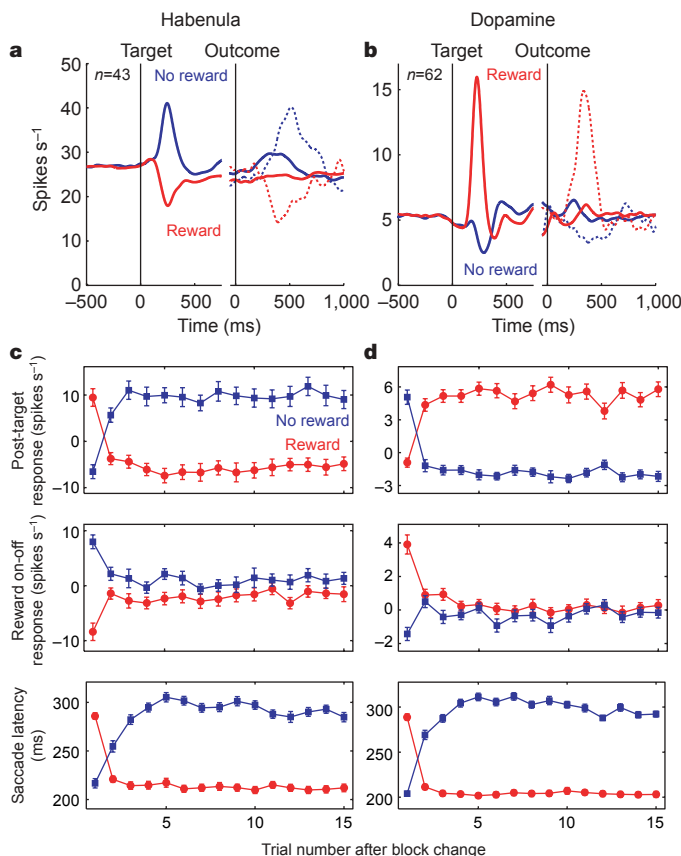


Figure 3 | Reward modulation of lateral habenula neurons and dopamine neurons. We compared population activity between lateral habenula neurons that showed a significant main effect of reward contingency in two-way ANOVA (**a** and **c**, $n = 43$) and dopamine neurons (**b** and **d**, $n = 62$). **a**, **b**, Averaged activity of habenula neurons (**a**) and dopamine neurons (**b**). Conventions are the same as SDFs in Fig. 2a. **c**, **d**, Changes in averaged post-target response (top), averaged reward on-off response (middle), and averaged saccade latency (bottom) after changes in position-reward contingency. Red circles and blue squares indicate the data in rewarded and unrewarded trials, respectively. In **a–d**, the data from monkey L and E as well as from ipsilateral and contralateral saccades are combined. Error bars in **c** and **d** indicate s.e.m.

Dopamine neurons also showed reward on-off responses in the first trials (dotted lines in the SDFs of Fig. 3b, and Supplementary Figs 5a and 6b) that declined rapidly after the transitions (Fig. 3d,

middle), but in the directions opposite to habenula neurons. Thus, the responses of both habenula and dopamine neurons shifted from the reward phase to the target phase as the prediction was established. These features suggest that habenula neurons encode reward prediction error, as dopamine neurons are thought to do¹. In support of this hypothesis, when the reward bias was smaller (that is, both targets were associated with rewards, but with different amounts), both habenula neurons and dopamine neurons responded to the targets differentially, but their responses were weaker (Supplementary Fig. 7 and Supplementary Note D). However, the responses of habenula neurons may not represent pure prediction error, as the reward on-off responses, albeit small, remained in habenula neurons (Fig. 3a, Fig. 3c middle); the difference in activity between unrewarded and rewarded trials remained significant ($P = 0.016$, Wilcoxon signed-rank test).

These results raise the possibility that habenula neuronal activity and dopamine neuronal activity are causally related. We found that, in unrewarded trials, the excitatory response of habenula neurons started earlier than the inhibitory response of dopamine neurons; in rewarded trials, however, the excitatory response of dopamine neurons started earlier than the inhibitory response of habenula neurons (Supplementary Note E). Thus, the excitation of habenula neurons could inhibit dopamine neurons in unrewarded trials, but inhibiting habenula neurons could not initiate the excitation of dopamine neurons in rewarded trials.

To test the hypothesis that habenula neurons affect dopamine reward responses, we delivered electrical stimulation (single biphasic pulse with 0.2 ms per phase duration, 100 μ A) to the lateral habenula during recording from 22 of the 62 dopamine neurons (15 in monkey L and 7 in monkey E). The dopamine neuron in Fig. 4a was strongly inhibited by the stimulation in the habenula on the same side. The averaged activity of the 22 dopamine neurons showed a significant inhibition during a time window from 10 to 40 ms after the onset of stimulation ($P < 0.05$, Wilcoxon signed-rank test) (filled circles in Fig. 4b). 18 of the 22 dopamine neurons (82%) were significantly inhibited during this window ($P < 0.01$, Wilcoxon signed-rank test). Electrical stimulation of either the ipsilateral or contralateral habenula was effective (ipsilateral, 9/10; contralateral, 9/12). Significant inhibitions were obtained with weaker stimulation, but the effects were more robust on the ipsilateral side: 75% of the dopamine neurons tested were significantly inhibited with 40 μ A (6/8; ipsilateral 4/4, contralateral 2/4); 50% with 20 μ A (6/12; ipsilateral 5/5, contralateral 1/7). In contrast, electrical stimulation of the surrounding thalamic area (mediodorsal thalamus, MD) was ineffective even when the stimulation sites were only 1 mm away from the lateral

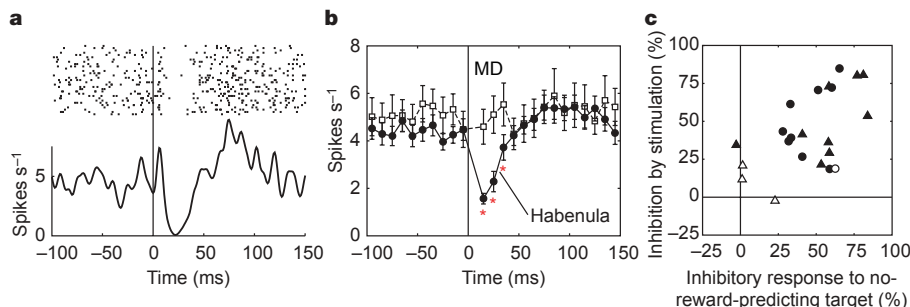


Figure 4 | Effects of electrical stimulation of the lateral habenula on dopamine neurons. **a**, Stimulation of the right lateral habenula (100 μ A, single biphasic pulse) induced a strong inhibition in a dopamine neuron in the right substantia nigra pars compacta. Rasters and SDF are aligned at the onset of electrical stimulation. Each row of rasters represents the timings of spikes combined across 10 trials. **b**, Averaged activity of dopamine neurons during stimulation of the lateral habenula (filled circles, $n = 22$) and during stimulation of the mediodorsal nucleus (MD) of the thalamus (open squares, $n = 8$). The averaged activity was calculated for each consecutive 10 ms bin. The data during 0–10 ms were excluded from the analysis (see Methods). Red asterisks indicate a significant inhibition compared with

background activity (–100 ms to 0 ms) ($P < 0.05$, Wilcoxon signed-rank test). Error bars indicate s.e.m. **c**, Comparison for each dopamine neuron between the inhibition by habenula stimulation (100 μ A) and the inhibitory response to the no-reward-predicting target ($n = 22$). The former was calculated during 10–40 ms after the onset of stimulation; the latter was calculated during 200–350 ms after the onset of the no-reward-predicting target. Circles and triangles indicate stimulation of the ipsilateral and contralateral lateral habenula, respectively. Filled symbols indicate neurons with statistically significant inhibitions by habenula stimulation ($P < 0.01$, Wilcoxon signed-rank test). Open symbols, no significance.

habenula: 8 dopamine neurons tested showed neither inhibition nor excitation ($P > 0.05$, Wilcoxon signed-rank test) (open squares in Fig. 4b).

The magnitude of the inhibition induced by habenula stimulation varied among the dopamine neurons. We found that dopamine neurons that are inhibited more strongly by habenula stimulation tended to show stronger inhibitions in response to the no-reward-predicting target ($r = 0.56$, $P < 0.01$) (Fig. 4c). The result is consistent with the hypothesis that the input from the lateral habenula underlies the phasic inhibition of dopamine neurons in response to the no-reward-predicting target. In contrast, the phasic excitation of dopamine neurons in response to the reward-predicting target was not correlated with the habenula-stimulation-induced inhibition (Supplementary Note F).

Using a saccade task, we have shown that lateral habenula neurons and dopamine neurons responded, with opposite signs, to visual targets that indicated the presence and absence of upcoming reward as well as to the unexpected delivery and omission of reward. The post-target responses of the habenula and dopamine neurons changed similarly after the reversal of the position-reward contingency, but in opposite directions. These results, together with the response latency analysis and the inhibition of dopamine neurons by electrical stimulation of the lateral habenula, suggest that the lateral habenula is capable of producing the negative reward response of dopamine neurons.

Recent studies from our laboratory have suggested that the reward modulation of dopamine neurons plays a key role in motivational control of saccadic eye movement^{22–24}. An underlying mechanism proposed from these studies is that the efficacy of cortico-caudate synapses carrying visuo-saccadic signals is enhanced or depressed depending on the concurrent increase or decrease, respectively, in dopaminergic inputs²³. According to this scheme, the lateral habenula would also be involved in the motivational control of saccadic eye movements by inhibiting dopamine neurons (Supplementary Fig. 1). In fact, the no-reward-dependent increase in the post-target response of habenula neurons was associated with the prolongation of saccade latency. However, testing this hypothesis would require further experiments, including artificial activation or inactivation of the lateral habenula.

One interpretation of our results may be that the lateral habenula is involved in negative reward processing while dopamine neurons are involved in positive reward processing. This view provides an interesting parallel with opponent-process theories, which postulate opponent interactions between an appetitive system and an aversive system^{25–27}. This hypothesis may be supported by previous studies showing that habenula neurons are activated while dopamine neurons are inhibited by aversive stimuli in anaesthetized rats^{14,28}. However, the nature of dopamine neuron has been studied under a wide variety of situations, including pavlovian and operant procedures²⁹. It remains to be determined whether habenula neurons show response patterns opposite to dopamine neurons in these situations as well.

Finally, it is unknown how lateral habenula neurons acquire the negative reward information. It might be provided by the inputs from the limbic areas³⁰ (Supplementary Fig. 2). The reward information might then be elaborated through the interplay among the lateral habenula, the basal ganglia, and monoaminergic (dopaminergic and serotonergic) systems. Our data suggest that the lateral habenula may play a pivotal role in the integrative function.

METHODS SUMMARY

Two adult rhesus monkeys (*Macaca mulatta*) were used for the experiments. All procedures for animal care and experimentation were approved by the Institute Animal Care and Use Committee and complied with the Public Health Service Policy on the humane care and use of laboratory animals. A plastic head holder and plastic recording chambers were fixed to the skull under general anaesthesia and sterile surgical conditions. Two search coils were surgically placed under the

conjunctiva of the eyes for recording of eye movements. We trained the monkeys to perform a one-direction-rewarded version of a visually guided saccade task (reward-biased visual saccade task)⁷ (Fig. 1a) and a control task (reward-biased memory saccade task)^{6,22} (Supplementary Fig. 4a). Details of these tasks can be found in Methods. While the monkeys were performing these tasks, we recorded single-unit activity from the lateral habenula and dopamine neurons in the substantia nigra pars compacta using conventional electrophysiological techniques (Methods). We estimated the positions of the lateral habenula and substantia nigra pars compacta by MRI. After the end of the recording session in one monkey, we confirmed the recording sites histologically (Supplementary Fig. 3 and Supplementary Note B). To examine the effects of lateral habenula neurons on dopamine neurons, we electrically stimulated the lateral habenula while recording from dopamine neurons. Details of the localization of the lateral habenula, identification of dopamine neurons, electrical stimulation, and histological procedures can be found in Methods. We analysed saccade latency and neuronal responses for trials in which the monkeys performed correct saccades. We focused on two kinds of neuronal responses: (1) post-target response that occurred after the onset of the saccade target, and (2) reward on-off response that occurred after the time when a reward was or would have been delivered. Further analysis methods can be found in Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 9 January; accepted 18 April 2007.

Published online 23 May 2007.

- Schultz, W. Predictive reward signal of dopamine neurons. *J. Neurophysiol.* **80**, 1–27 (1998).
- Houk, J. C., Adams, J. L. & Barto, A. in *Models of Information Processing in the Basal Ganglia* (eds Houk, J. C., Davis, J. L. & Beiser, D. G.) 249–270 (MIT Press, Cambridge, Massachusetts, 1995).
- Schultz, W., Dayan, P. & Montague, P. R. A neural substrate of prediction and reward. *Science* **275**, 1593–1599 (1997).
- Montague, P. R., Dayan, P. & Sejnowski, T. J. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J. Neurosci.* **16**, 1936–1947 (1996).
- Satoh, T., Nakai, S., Sato, T. & Kimura, M. Correlated coding of motivation and outcome of decision by dopamine neurons. *J. Neurosci.* **23**, 9913–9923 (2003).
- Takikawa, Y., Kawagoe, R. & Hikosaka, O. A possible role of midbrain dopamine neurons in short- and long-term adaptation of saccades to position-reward mapping. *J. Neurophysiol.* **92**, 2520–2529 (2004).
- Lauwereyns, J., Watanabe, K., Coe, B. & Hikosaka, O. A neural correlate of response bias in monkey caudate nucleus. *Nature* **418**, 413–417 (2002).
- Nakahara, H., Itoh, H., Kawagoe, R., Takikawa, Y. & Hikosaka, O. Dopamine neurons can represent context-dependent prediction error. *Neuron* **41**, 269–280 (2004).
- Herkenham, M. & Nauta, W. J. Efferent connections of the habenular nuclei in the rat. *J. Comp. Neurol.* **187**, 19–47 (1979).
- Christoph, G. R., Leonzio, R. J. & Wilcox, K. S. Stimulation of the lateral habenula inhibits dopamine-containing neurons in the substantia nigra and ventral tegmental area of the rat. *J. Neurosci.* **6**, 613–619 (1986).
- Murphy, C. A., DiCamillo, A. M., Haun, F. & Murray, M. Lesion of the habenular efferent pathway produces anxiety and locomotor hyperactivity in rats: A comparison of the effects of neonatal and adult lesions. *Behav. Brain Res.* **81**, 43–52 (1996).
- Amat, J. et al. The role of the habenular complex in the elevation of dorsal raphe nucleus serotonin and the changes in the behavioral responses produced by uncontrollable stress. *Brain Res.* **917**, 118–126 (2001).
- Caldecott-Hazard, S., Mazziotto, J. & Phelps, M. Cerebral correlates of depressed behavior in rats, visualized using ¹⁴C-2-deoxyglucose autoradiography. *J. Neurosci.* **8**, 1951–1961 (1988).
- Gao, D. M., Hoffman, D. & Benabid, A. L. Simultaneous recording of spontaneous activities and nociceptive responses from neurons in the pars compacta of substantia nigra and in the lateral habenula. *Eur. J. Neurosci.* **8**, 1474–1478 (1996).
- Thornton, E. W. & Bradbury, G. E. Effort and stress influence the effect of lesion of the habenula complex in one-way active avoidance learning. *Physiol. Behav.* **45**, 929–935 (1989).
- Sutherland, R. J. The dorsal diencephalic conduction system: A review of the anatomy and functions of the habenular complex. *Neurosci. Biobehav. Rev.* **6**, 1–13 (1982).
- Lecourtier, L. & Kelly, P. H. Bilateral lesions of the habenula induce attentional disturbances in rats. *Neuropsychopharmacology* **30**, 484–496 (2005).
- Ullsperger, M. & von Cramon, D. Y. Error monitoring using external feedback: Specific roles of the habenular complex, the reward system, and the cingulate motor area revealed by functional magnetic resonance imaging. *J. Neurosci.* **23**, 4308–4314 (2003).

19. Shepard, P. D., Holcomb, H. H. & Gold, J. M. Schizophrenia in translation: the presence of absence: habenular regulation of dopamine neurons and the encoding of negative outcomes. *Schizophr. Bull.* **32**, 417–421 (2006).
20. Ellison, G. Stimulant-induced psychosis, the dopamine theory of schizophrenia, and the habenula. *Brain Res. Brain Res. Rev.* **19**, 223–239 (1994).
21. Sandyk, R. Relevance of the habenular complex to neuropsychiatry: A review and hypothesis. *Int. J. Neurosci.* **61**, 189–219 (1991).
22. Kawagoe, R., Takikawa, Y. & Hikosaka, O. Reward-predicting activity of dopamine and caudate neurons—a possible mechanism of motivational control of saccadic eye movement. *J. Neurophysiol.* **91**, 1013–1024 (2004).
23. Hikosaka, O., Nakamura, K. & Nakahara, H. Basal ganglia orient eyes to reward. *J. Neurophysiol.* **95**, 567–584 (2006).
24. Nakamura, K. & Hikosaka, O. Role of dopamine in the primate caudate nucleus in reward modulation of saccades. *J. Neurosci.* **26**, 5360–5369 (2006).
25. Daw, N. D., Kakade, S. & Dayan, P. Opponent interactions between serotonin and dopamine. *Neural Netw.* **15**, 603–616 (2002).
26. Seymour, B. *et al.* Opponent appetitive-aversive neural processes underlie predictive learning of pain relief. *Nature Neurosci.* **8**, 1234–1240 (2005).
27. Solomon, R. L. & Corbit, J. D. An opponent-process theory of motivation. I. Temporal dynamics of affect. *Psychol. Rev.* **81**, 119–145 (1974).
28. Ungless, M. A., Magill, P. J. & Bolam, J. P. Uniform inhibition of dopamine neurons in the ventral tegmental area by aversive stimuli. *Science* **303**, 2040–2042 (2004).
29. Schultz, W. Getting formal with dopamine and reward. *Neuron* **36**, 241–263 (2002).
30. Parent, A., Gravel, S. & Boucher, R. The origin of forebrain afferents to the habenula in rat, cat and monkey. *Brain Res. Bull.* **6**, 23–38 (1981).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. H. Wurtz, B. J. Richmond, K. Nakamura, L. Ding, M. Isoda, S. Hong and E. Bromberg-Martin for discussions, M. K. Smith for histological expertise, and J. W. McClurkin, A. M. Nichols, T. W. Ruffner, A. V. Hays and L. P. Jensen for technical assistance. This research was supported by the Intramural Research Program at the National Institutes of Health, National Eye Institute.

Author Contributions M.M. performed the experiments and analysed the data. M.M. and O.H. discussed the results and wrote the manuscript. O.H. organized this project.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to O.H. (oh@lsr.nei.nih.gov).

METHODS

Behavioural task. Behavioural tasks were under the control of a QNX-based real-time experimentation data acquisition system (REX, Laboratory of Sensorimotor Research, National Eye Institute, National Institutes of Health (LSR/NEI/NIH), Bethesda, Maryland). The monkeys sat in a primate chair, facing a frontoparallel screen 33 cm from the monkey's eyes in a sound-attenuated and electrically shielded room. Stimuli generated by an active matrix liquid crystal display projector (PJ550, ViewSonic) were rear-projected on the screen. The monkeys were trained to perform a one-direction-rewarded version of the visually guided saccade task⁷ (Fig. 1a), which we call 'reward-biased visual saccade task'. A trial started when a small fixation spot appeared on the screen. After the monkeys maintained fixation on the spot for 1,200 ms, the fixation spot disappeared and a peripheral target appeared at either right or left, 20° (for monkey L) or 15° (for monkey E) from the fixation spot. The monkeys were required to make a saccade to the target within 500 ms. Correct and incorrect saccades were signalled by tone and beep stimuli 200 ms after the saccades. Within a block of 24 trials, saccades to one fixed direction were rewarded with 0.3 ml of apple juice while saccades to the other direction were not rewarded. The position-reward contingency was reversed in the next block with no external instruction. Even in the unrewarded trials, the monkeys had to make a correct saccade; otherwise, the same trial was repeated. In rewarded trials a liquid reward was delivered which started simultaneously with the tone stimulus.

As a control, the monkeys were also trained to perform a reward-biased memory saccade task^{6,22} (Supplementary Fig. 4a). After the monkeys maintained fixation for 1,200 ms, a peripheral target was presented for 100 ms at either right or left, 20° (for monkey L) or 15° (for monkey E) from the fixation spot. The monkeys had to maintain fixation while remembering the target position. After 1,000 ms (for monkey L) or 800 ms (for monkey E) delay, the fixation spot disappeared, and the monkeys were required to make a saccade to the remembered target position. Other procedures including the position-reward contingency were the same as the reward-biased visual saccade task.

Electrophysiology. One recording chamber was placed over the midline of the parietal cortex, tilted posteriorly by 38°, and was aimed at the habenula; the other recording chamber was placed over the fronto-parietal cortex, tilted laterally by 35°, and was aimed at the substantia nigra. Single-unit recordings and electrical stimulations were performed using tungsten electrodes (Frederick Haer) that were advanced by an oil-driven micro-manipulator (MO-97A, Narishige) or an electrically driven micro-manipulator (MicroStepper, LSR/NEI/NIH). The recording and stimulation sites were determined using a grid system, which allowed recordings at every 1 mm between penetrations. The electrode was introduced into the brain through a stainless steel guide tube, which was inserted into one of the grid holes and then to the brain via the dura. For finer mapping of neurons, we also used a complementary grid, which allowed electrode penetrations between the holes of the original grid. Single neurons were isolated on-line using a custom voltage-time window discrimination software (MEX, LSR/NEI/NIH).

Localization of the lateral habenula. We estimated the position of the habenula by obtaining MRIs (4.7T, Bruker). We then recorded from neurons in and around the estimated habenula, and found that the firing patterns and spike shapes within the estimated habenula were distinctly different from neurons in the surrounding thalamic area (mediodorsal thalamus, MD). Presumed habenula neurons fired tonically with relatively high background rates (mean \pm s.d. = 27.8 ± 14.5 spikes s^{-1} , $n = 49$). In contrast, presumed MD neurons exhibited irregular and bursty firing with lower background rates (mean \pm s.d. = 7.5 ± 5.4 spikes s^{-1} , $n = 33$) and their action potentials were much broader than those of habenula neurons. Furthermore, most of the presumed habenula neurons, but none of the presumed MD neurons, were sensitive to reward outcome. Presumed habenula neurons were recorded at penetrations 1.5 or 2.0 mm from the midline with 1 or 2 penetrations separated anteroposteriorly; they were recorded at 2 holes (separated by 1 mm) at most in each hemisphere for a given grid. We made only two penetrations at 1.0 mm from the midline, but the recorded neurons were not sensitive to reward outcome; they were judged to be within the medial habenula. Recordings at penetrations 1 mm away from the presumed lateral habenula laterally or anteriorly yielded presumed MD neurons. Recordings at 1 mm posteriorly were different from others in that the first neuron in the subcortical structure was considerably deeper and neurons there may respond to visual stimuli but not reward outcome, suggesting that the electrodes were in the pretectum. Importantly, the characteristics of firing and the relation to reward outcome were distinctly different between the presumed habenula and the presumed MD or the pretectum, even when they were separated only by 0.5 mm or 1.0 mm. This was also true for the effect of electrical stimulation (see Fig. 4).

In the present study we isolated stable action-potentials from 74 habenula neurons. For each of these neurons, we first examined its activity using the

reward-biased visual saccade task without recording. If we found any task-related response on-line, we recorded the activity of the neuron using the task. If not, we did not examine the activity further. Of the 74 habenula neurons, 49 were regarded as task-related neurons. The activity of these 49 habenula neurons was recorded and comprised the sample used for the analysis. However, we are not completely confident whether all neurons in the lateral habenula can be characterized as presented in this study. In particular, there may be different types of neurons in the deeper part of the lateral habenula that we have not explored fully.

Identification of dopamine neurons. We searched for dopamine neurons in and around the substantia nigra pars compacta. Dopamine neurons were identified by their irregular and tonic firing around 5 spikes s^{-1} with broad spike potentials. In this experiment, we focused on dopamine neurons that responded to reward-predicting stimuli with a phasic excitation, and recorded from 62 dopamine neurons. Dopamine-like neurons that were not sensitive to reward-predicting stimuli were not examined further.

Electrical stimulation. In order to examine the effect of electrical stimulation in the lateral habenula on the activity of dopamine neurons, we first recorded single- or multi-unit activity in the habenula that was modulated by reward outcome in the reward-biased visual saccade task, and then used the electrode for electrical stimulation. We then recorded from a dopamine neuron that responded to reward-predicting stimuli with a phasic excitation from the substantia nigra chamber, and delivered a single current pulse (biphasic negative-positive pulse with 0.2 ms per phase duration) through the habenula electrode. The default setting of stimulation current was 100 μA . If the dopamine neuron was inhibited by 100 μA stimulation, we also used 20 and/or 40 μA .

Data analysis. We defined the post-target response as the discharge rate during 150–350 ms after the target onset minus the background discharge rate before the target onset (500–0 ms). The reward on-off response was defined as the discharge rate during 250–700 ms after the onset of the tone stimulus (which was synchronized with reward onset if reward was present) minus the background discharge rate. These time windows for post-target and reward on-off responses were determined on the basis of the averaged activity of habenula neurons and that of dopamine neurons. Specifically, we set the time windows such that they include major parts of the excitatory and inhibitory responses of both habenula and dopamine neurons.

To evaluate the relative contribution of reward contingency (reward or no-reward) and target position to the post-target response, we performed a two-way ANOVA and calculated reward index and position index for each neuron. The reward and position indices were defined as the percentage of variance accounted for by reward contingency and by target position, respectively³¹.

The latency of the averaged post-target response was calculated for each of four conditions (ipsilateral reward, ipsilateral no-reward, contralateral reward, and contralateral no-reward) using a bootstrap analysis. The data set of each neuron consisted of at least 24 trials for each condition. For each neuron, the trials were randomly resampled with replacements to form a new bootstrap data set which had the same number of trials as the original data set. The bootstrap data sets of all neurons were combined, and at each time point after target onset their averaged discharge rate was calculated during the 25 ms period before the time point (pre-period) and during the 25 ms period after the time point (post-period). Then, the average discharge rate during the pre-period was compared with that during the post-period. Such random resampling and comparison were repeated 1,000 times. If the averaged discharge rate was larger during the post-period than during the pre-period in >975 repetitions, the time point was regarded as a time of significant increase. In contrast, if the averaged discharge rate was smaller during the post-period than during the pre-period in >975 repetitions, the time point was regarded as a time of significant decrease. This procedure was repeated by shifting the time point in 1 ms steps after target onset. If, of 20 consecutive 1 ms steps, the beginning and at least 19 showed significant increases or decreases, the beginning was defined as the latency of excitatory or inhibitory responses, respectively.

For the control task (reward-biased memory saccade task), we analysed the response just after the target onset (the discharge rate during 150–350 ms after the target onset minus the background discharge rate) and the response around the saccade onset (the discharge rate during –100 to 100 ms after the saccade onset minus the background discharge rate).

In the population analysis of the electrical stimulation experiment (Fig. 4b), we excluded the activity during 0–10 ms after the stimulation onset because electrical stimulation often generated brief electric noise which contaminated the electrophysiological recording.

Histology. After the end of the recording session in monkey L, we selected representative locations for electrode penetrations into the lateral habenula, substantia nigra pars compacta and MD. When typical single- or multi-unit activities were recorded for each region, we made electrolytic microlesions at

the recording sites (20 μ A and 20 s). Then monkey L was deeply anaesthetized with an overdose of pentobarbital sodium, and perfused with 10% formaldehyde. The brain was blocked and equilibrated with 10% sucrose. Frozen sections were cut every 50 μ m in the plane parallel to the electrode penetration into the habenula. The sections were stained with cresyl violet.

31. Paton, J. J., Belova, M. A., Morrison, S. E. & Salzman, C. D. The primate amygdala represents the positive and negative value of visual stimuli during learning. *Nature* **439**, 865–870 (2006).

LETTERS

Macrophage-specific PPAR γ controls alternative activation and improves insulin resistance

Justin I. Odegaard^{1,2*}, Roberto R. Ricardo-Gonzalez^{1,2*}, Matthew H. Goforth¹, Christine R. Morel¹, Vidya Subramanian⁴, Lata Mukundan¹, Alex Red Eagle^{1,3}, Divya Vats¹, Frank Brombacher⁵, Anthony W. Ferrante⁴ & Ajay Chawla^{1,2}

Obesity and insulin resistance, the cardinal features of metabolic syndrome, are closely associated with a state of low-grade inflammation^{1,2}. In adipose tissue chronic overnutrition leads to macrophage infiltration, resulting in local inflammation that potentiates insulin resistance^{3,4}. For instance, transgenic expression of *Mcp1* (also known as chemokine ligand 2, *Ccl2*) in adipose tissue increases macrophage infiltration, inflammation and insulin resistance^{5,6}. Conversely, disruption of *Mcp1* or its receptor *Ccr2* impairs migration of macrophages into adipose tissue, thereby lowering adipose tissue inflammation and improving insulin sensitivity^{5,7}. These findings together suggest a correlation between macrophage content in adipose tissue and insulin resistance. However, resident macrophages in tissues display tremendous heterogeneity in their activities and functions, primarily reflecting their local metabolic and immune microenvironment⁸. While *Mcp1* directs recruitment of pro-inflammatory classically activated macrophages to sites of tissue damage^{5,8}, resident macrophages, such as those present in the adipose tissue of lean mice, display the alternatively activated phenotype⁹. Despite their higher capacity to repair tissue¹⁰, the precise role of alternatively activated macrophages in obesity-induced insulin resistance remains unknown. Using mice with macrophage-specific deletion of the peroxisome proliferator activated receptor- γ (PPAR γ), we show here that PPAR γ is required for maturation of alternatively activated macrophages. Disruption of PPAR γ in myeloid cells impairs alternative macrophage activation, and predisposes these animals to development of diet-induced obesity, insulin resistance, and glucose intolerance. Furthermore, gene expression profiling revealed that downregulation of oxidative phosphorylation gene expression in skeletal muscle and liver leads to decreased insulin sensitivity in these tissues. Together, our findings suggest that resident alternatively activated macrophages have a beneficial role in regulating nutrient homeostasis and suggest that macrophage polarization towards the alternative state might be a useful strategy for treating type 2 diabetes.

To distinguish between the pathogenic and reparative functions of macrophages in metabolic disease, we sought to identify transcriptional regulators that control alternative macrophage activation. Notably, PPAR γ , a genetic sensor of fatty acids¹¹, is markedly induced in macrophages stimulated with interleukin-4 (IL-4)^{12,13}, prompting us to examine its role in alternative macrophage activation. Because Balb/c mice—but not C57Bl/6 mice—can fully support maturation of alternatively activated macrophages¹⁴, we generated macrophage-specific PPAR γ knockout (Mac-PPAR γ KO) mice on the T helper 2 (Th2)-permissive Balb/c strain (see Methods). Quantitative

polymerase chain reaction (Q-PCR) analysis of genomic DNA from peritoneal elicited cells and bone-marrow-derived macrophages showed high excision efficiency (~85–90%) in Mac-PPAR γ KO animals (Supplementary Fig. 1a). Immunoblot analysis further confirmed the absence of PPAR γ protein in bone-marrow-derived macrophages from Mac-PPAR γ KO mice (Supplementary Fig. 1b). Verifying a critical role for PPAR γ in alternative activation, arginase I messenger RNA and activity—both hallmarks of alternatively activated macrophages¹⁰—were reduced by 40% and 50%, respectively, in IL-4 stimulated PPAR γ null bone-marrow-derived macrophages (Supplementary Fig. 1c and Fig. 1a).

To determine whether the arginase I gene is a direct transcriptional target of PPAR/RXR heterodimers, we analysed its promoter region for PPAR response elements. A putative PPAR response element was identified in the distal enhancer of the arginase I gene, a region known to be essential for its expression in alternatively activated macrophages¹⁵. Electromobility gel shift assays confirmed that PPAR/RXR heterodimers bound to the identified site in a sequence-specific manner (Supplementary Fig. 1d). To verify that PPAR/RXR heterodimers can activate the arginase I promoter *in vivo*, cells of the mouse macrophage cell line RAW264.7 were transiently transfected with luciferase reporter construct driven by ~4 kilobases (kb) of mouse arginase I promoter/enhancer. Treatment of transfected macrophages with IL-4 led to a ~9-fold increase in luciferase activity (Fig. 1b). Moreover, the addition of PPAR γ and its ligand, rosiglitazone, enhanced the ability of IL-4 to activate the arginase I promoter (twice the IL-4 stimulated level), demonstrating that PPAR γ directly regulates this important facet of macrophage activation. Because alternatively activated macrophages can counteract excessive secretion of pro-inflammatory cytokines¹³, we examined whether IL-4 could appropriately attenuate lipopolysaccharide-induced TNF α and IL-6 secretion. Although lipopolysaccharide-stimulated release of TNF α and IL-6 was not significantly different between the two genotypes (Fig. 1c and Supplementary Fig. 1e), IL-4 failed to suppress the secretion of IL-6 in macrophages deficient in PPAR γ (Fig. 1c), which indicates that a subset of IL-4 dependent anti-inflammatory responses are regulated by PPAR γ .

We recently reported that a switch to oxidative metabolism is an integral component of alternative macrophage activation¹³. Because PPARs regulate fatty acid homeostasis in many cell types¹⁶, we investigated the requirement for PPAR γ or PPAR δ (the two major PPAR subtypes expressed in murine macrophages) in controlling this oxidative switch. Surprisingly, PPAR γ , rather than PPAR δ , was required for IL-4 induced increase in β -oxidation of fatty acids, as shown by the

¹Division of Endocrinology, Metabolism and Gerontology, Department of Medicine. ²Graduate Program in Immunology. ³Department of Genetics, Stanford University School of Medicine, Stanford, California 94305-5103, USA. ⁴Department of Medicine, Naomi Berrie Diabetes Center, Columbia University College of Physicians and Surgeons, New York, New York 10032, USA. ⁵Institute of Infectious Disease and Molecular Medicine, Division of Immunology, Health Science Faculty, University of Cape Town, Werhner Beit South, Observatory, 7925, Anzioroad, Cape Town, South Africa.

*These authors contributed equally to this work.

~70% reduction in the rate of fatty acid oxidation in IL-4-stimulated PPAR γ null macrophages (Fig. 1d). In agreement with this genetic data, ligand activation of PPAR γ led to a doubling of the rate of β -oxidation (Supplementary Fig. 1f). Moreover, Q-PCR analyses with reverse transcription revealed that expression of genes controlling lipolysis (*Lpl*), fatty acid uptake (*CD36*), and oxidation (*Acadm*, *Acadl*) was reduced in IL-4-stimulated PPAR γ null macrophages (Supplementary Fig. 1g). In contrast, fatty acid uptake rates were similar in control and Mac-PPAR γ KO macrophages (Supplementary Fig. 1h).

An increase in cellular mitochondrial content accompanies this metabolic switch, so we tested the requirement for PPAR γ in mitochondrial biogenesis. Cellular staining with MitoTracker Green revealed an absolute requirement for PPAR γ in mediating the biogenic effects of IL-4 (Fig. 1e). Fluorescence microscopy with MitoTracker Red further verified that reduction in respiring mitochondria largely accounted for the observed decrease in total mitochondria in these cells

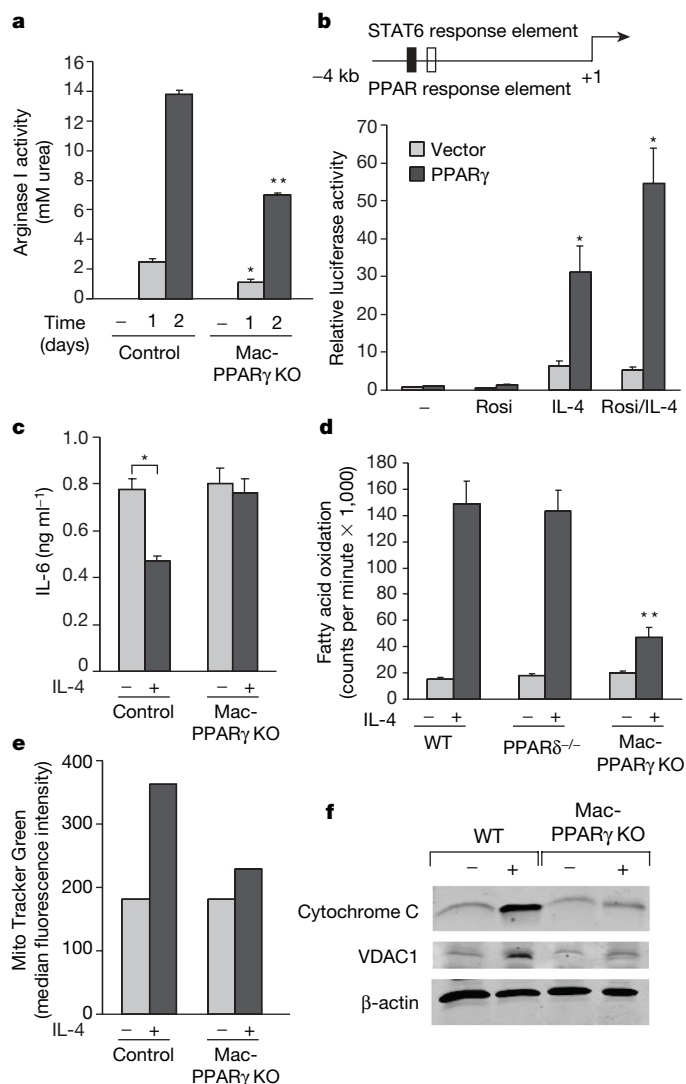


Figure 1 | PPAR γ regulates alternative macrophage activation. **a**, Decreased induction of arginase I activity by IL-4 in PPAR γ null macrophages. **b**, Activation of arginase I promoter by PPAR γ /RXR heterodimers. Rosi, rosiglitazone. **c**, PPAR γ is required for suppression of IL-6 production in alternatively activated macrophages. **d**, PPAR γ is required for macrophage oxidative metabolism. Fatty acid oxidation rates were quantified in control, PPAR δ null and PPAR γ null bone-marrow-derived macrophages 96 h after stimulation with IL-4. **e**, **f**, IL-4 fails to induce mitochondrial biogenesis in PPAR γ -deficient macrophages, as measured by Mito Tracker Green (**e**) and cytochrome C and VDAC1 (**f**) protein levels. Equivalent loading was confirmed by immunoblotting for β -actin. Data presented as mean \pm s.e.m. * P < 0.05; ** P < 0.01.

(Supplementary Fig. 1i). Consistent with these findings, IL-4 failed to induce the mitochondrial proteins cytochrome C and voltage-dependent anion channel 1 (VDAC1) in PPAR γ -deficient macrophages (Fig. 1f). Lastly, the observed impairment in alternative activation was independent of the upstream IL-4 signalling pathway, because PPAR γ -deficient macrophages expressed similar levels of IL-4R α protein on their cell surface and stimulation with IL-4 resulted in equivalent levels of phosphorylation of the transcription factor STAT6 (Supplementary Fig. 2a, b).

To provide an independent test for whether PPAR γ can regulate the macrophage program of alternative activation *in vivo*, we used an immunological model for studying macrophage Th2-type responses. Cutaneous infection with *Leishmania major* in Th2-biased Balb/c mice leads to non-healing lesions, whereas the Th1-prone C57Bl/6 mice are resistant to acute infection¹⁷. Importantly, because impairment in alternative macrophage activation can delay disease progression in Balb/c mice¹⁸, we investigated whether Mac-PPAR γ KO mice were also resistant to cutaneous leishmaniasis. Indeed, Mac-PPAR γ KO mice had significantly less footpad swelling 5–7 weeks after injection of *L. major* promastigotes (Fig. 2a). Lesions in Mac-PPAR γ KO started to stabilize after 7 weeks, but the footpads of control mice continued to enlarge and rapidly underwent necrosis (Fig. 2b). Consistent with local extension of disease, the draining popliteal lymph nodes were hypercellular and enlarged in control mice (Supplementary Fig. 3). Because these results mimic the phenotype of mice lacking alternatively activated macrophages, our data strongly suggest that PPAR γ is required for acquisition and maintenance of the alternatively activated phenotype.

On the basis of these results, we tested whether genetic deletion of PPAR γ in macrophages would exacerbate the development of metabolic syndrome. Cohorts of control and Mac-PPAR γ KO mice were challenged with a high fat diet for 18 weeks to promote maximal infiltration of macrophages into adipose tissue⁴. Surprisingly, Mac-PPAR γ KO mice gained more weight than control mice on the high fat diet. After 17 weeks on a high fat diet, the body weight of Mac-PPAR γ KO mice (46.7 ± 2.3 g) exceeded that of control mice (40.3 ± 1.2 g) by ~15% (Fig. 3a). Dual-energy X-ray absorptiometry showed a 20% increase in total fat mass and a 12% increase in adiposity in Mac-PPAR γ KO animals (Fig. 3b). Consistent with an increase in adiposity, epididymal fat pads were larger (~18% by mass) and serum leptin levels were higher (~57%) in Mac-PPAR γ KO mice (Supplementary Fig. 4a, b)¹⁹. However, adipocyte cell size was not significantly different (Supplementary Fig. 4c, d).

To define the alterations in adipose tissue better, we quantified the transcript levels of genes important in adipocyte differentiation and fatty acid metabolism. As shown in Fig. 3c, mRNA levels of a large number of genes involved in nutrient uptake, fatty acid synthesis, and β -oxidation were reduced by ~50–80% in the white adipose tissue of Mac-PPAR γ KO mice, indicating a global suppression of adipocyte function. Direct co-culture of PPAR γ -deficient macrophages with adipocytes led to a marked reduction of insulin-stimulated glucose

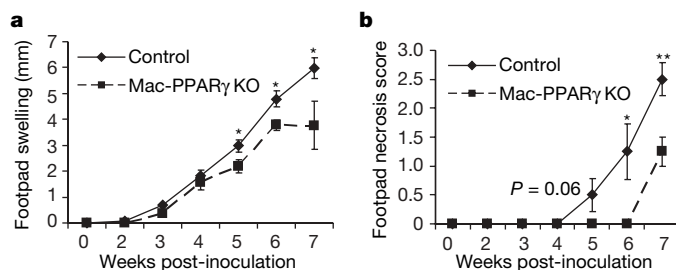


Figure 2 | Mac-PPAR γ KO mice are less susceptible to infection by *Leishmania major*. **a**, Footpad swelling in control and Mac-PPAR γ KO mice after infection with *L. major* (n = 5 per genotype). **b**, Decreased necrosis in footpads of Mac-PPAR γ KO mice. Data presented as mean \pm s.e.m. * P < 0.05; ** P < 0.01.

uptake in adipocytes (Fig. 3d), suggesting a causal relationship between PPAR γ -deficient adipose tissue macrophages (ATMs) and dysregulation of adipocyte metabolism. Remarkably, this was confirmed by Q-PCR analysis of co-cultured adipocytes, which showed suppression of adipocyte gene expression by PPAR γ -deficient macrophages (Supplementary Fig. 4e).

To determine whether macrophage infiltration or aberrant activation contributed to adiposity, we analysed macrophage-specific gene expression in the white adipose tissue of control and Mac-PPAR γ KO mice. Despite being more obese, transcript levels of macrophage-specific markers, *Emr1* and *CD68*, were reduced by ~70% in the white adipose tissue of Mac-PPAR γ KO mice (Fig. 3e). Consistent with the Q-PCR data, histological analysis showed marked reduction in the macrophage content of the white adipose tissue of Mac-PPAR γ KO mice (Fig. 3f, g). Importantly, expression of genes preferentially expressed in alternatively activated macrophages¹⁰, such as *Arg1*, *Mrc1*, and *Clec7a*, was also decreased by ~70–80% in the white adipose tissue of Mac-PPAR γ KO mice (Fig. 3e), suggesting that macrophage-specific deletion of PPAR γ leads to specific reduction

in alternatively activated macrophages. The absence of alternatively activated ATMs increased local inflammation in white adipose tissue, as shown by higher expression of *IL-6* and *Nos2* (Fig. 3e). Lastly, although ATM content is much lower in lean mice^{3,9}, Q-PCR analysis of white adipose tissue revealed that chow-fed (a low-fat diet) Mac-PPAR γ KO mice expressed much lower levels of genes associated with the alternative state (Supplementary Fig. S5a).

While the Th2 cytokines IL-4 and IL-13 are required for maturation of alternatively activated macrophages during parasitic infections^{10,14}, the importance of this signalling pathway in ATM biology has not been formally tested. To determine whether IL-4/IL-13 signalling is required for acquisition of the alternative phenotype by ATMs, we analysed the molecular signature of ATMs in mice defective in IL-4/IL-13 signalling, in particular, the STAT6 null and Mac-IL-4R α KO mice^{14,20}. Consistent with a critical role for Th2 cytokines in alternative activation, the molecular signature of alternatively activated ATMs was dramatically reduced in lean STAT6 null and Mac-IL-4R α KO mice (Supplementary Fig. 5b, c). Moreover, the expression profile of ATMs in obese STAT6 null mice was identical to

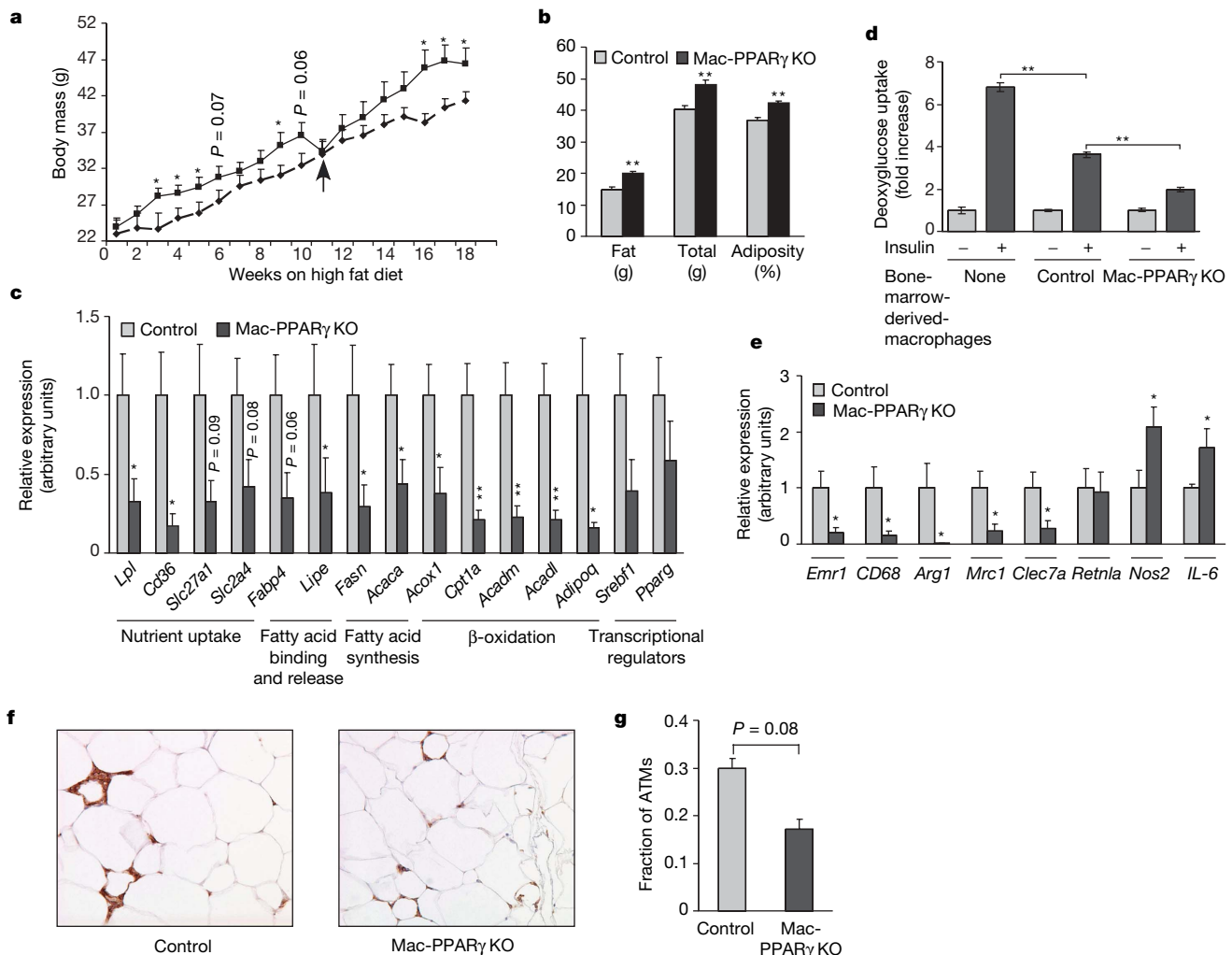


Figure 3 | Alterations in adipose tissue mass and function in Mac-PPAR γ KO mice. **a**, Weight gain of control and Mac-PPAR γ KO mice on a high fat diet. Arrow indicates mice fasted for glucose and insulin tolerance tests. **b**, Body composition as determined by dual-energy X-ray absorptiometry ($n = 5$ per genotype). **c**, Q-PCR analyses of gonadal adipose tissue gene expression. Relative transcript levels of genes involved in adipocyte differentiation and function. *Lpl*, lipoprotein lipase; *Cd36*, fatty acid translocase; *Slc27a1*, fatty acid transporter 1; *Slc2a4*, glucose transporter 4; *Fabp4*, fatty acid binding protein 4; *Lipe*, hormone sensitive lipase; *Fasn*, fatty acid synthase; *Acaca*, acetyl-Coenzyme A carboxylase; *Acox1*, acyl-Coenzyme A oxidase 1; *Cpt1a*, carnitine palmitoyltransferase 1a; *Acadm* and *Acadl*, medium- and long-chain

acyl-CoA dehydrogenase; *Adipoq*, adiponectin; *Srebf1*, sterol regulatory element binding factor 1; *Pparg*, peroxisome proliferator activated receptor γ . **d**, Co-culture of macrophages with adipocytes decreases insulin-stimulated glucose uptake. **e**, Q-PCR analyses of macrophage gene expression in white adipose tissue from control and Mac-PPAR γ KO mice. *Emr1*, F4/80; *Cd68*, macrophage; *Arg1*, arginase I; *Mrc1*, mannose receptor; *Clec7a*, dectin-1; *Retnla*, resistin-like α ; *Nos2*, inducible nitric oxide synthase; *IL-6*, interleukin-6. **f**, **g**, Macrophage content of white adipose tissue as assessed by F4/80 staining (**f**). Fraction of ATMs is equal to F4/80-stained cells/total cells counted in the fields (**g**). Data was statistically analysed using the paired Student's *t*-test (**g**). Data presented as mean \pm s.e.m. * $P < 0.05$; ** $P < 0.01$.

the profile observed in obese Mac-PPAR γ KO mice (compare Fig. 3e and Supplementary Fig. 5d). The molecular signatures of ATMs in three distinct genetic models (Mac-PPAR γ KO, STAT6 null and Mac-IL-4R α KO) were similar, so our data show that an intact IL-4/IL-13/STAT6/PPAR γ axis is required for maturation of alternatively activated ATMs. Furthermore, to investigate whether stimulation of ATMs by IL-4 is sufficient to polarize them towards the alternative state, obese control and Mac-PPAR γ KO mice were injected with recombinant IL-4, and ATM gene expression was monitored by Q-PCR. Supplementary Fig. 5e shows that IL-4 induced the expression of alternatively activated markers (*Arg1*, *Chi3l3*, *Mrc1* and *Jag1*) in a PPAR γ -dependent manner.

To explore the role of alternatively activated macrophages in obesity-induced insulin resistance, we performed glucose and insulin tolerance tests in control and Mac-PPAR γ KO mice. Oral glucose tolerance tests revealed that Mac-PPAR γ KO mice were significantly more glucose intolerant after an 18-week high fat diet challenge (Fig. 4a). As would be expected with a decrease in insulin sensitivity, Mac-PPAR γ KO mice were more resistant to the glucose lowering effects of exogenous insulin (Fig. 4b). Moreover, after a 4-h fast, paired blood glucose and serum insulin levels were much higher in Mac-PPAR γ KO mice, with

blood glucose levels at 170 ± 9.2 mg dl $^{-1}$ versus 125 ± 2.3 mg dl $^{-1}$ ($P < 0.0007$) and serum insulin levels at $(3.00 \pm 0.77$ ng ml $^{-1}$ versus 0.24 ± 0.02 ng ml $^{-1}$ ($P < 0.002$), (Fig. 4c). The calculated homeostasis model assessment (HOMA) measure of insulin resistance was significantly higher in Mac-PPAR γ KO mice (Fig. 4d). To investigate the potential sites of insulin resistance, control and Mac-PPAR γ KO mice were injected with saline or insulin (5 mU g $^{-1}$) through their inferior vena cava, and liver and skeletal muscle were quickly harvested for biochemical analysis. Strikingly, insulin-stimulated phosphorylation of AKT was markedly decreased in liver and skeletal muscle of Mac-PPAR γ KO mice (Fig. 4e, f), findings consistent with the presence of insulin resistance in these tissues.

Mitochondrial dysfunction in muscle is associated with the onset of insulin resistance and type 2 diabetes^{21–24}, prompting us to examine mitochondrial gene expression in the skeletal muscles of obese mice. Q-PCR analyses showed that mRNAs encoding key enzymes in fatty acid oxidation (*Cpt1b*, *Acox1*) and oxidative phosphorylation (*Ndufs1*, *Sdh*, *Atp5j* and *Atp5b*) were reduced by 30–70% in the quadriceps muscles of Mac-PPAR γ KO mice (Fig. 4g). Moreover, expression level of transcription factors and coactivator proteins controlling mitochondrial biogenesis^{25,26}, including *Tfam*,

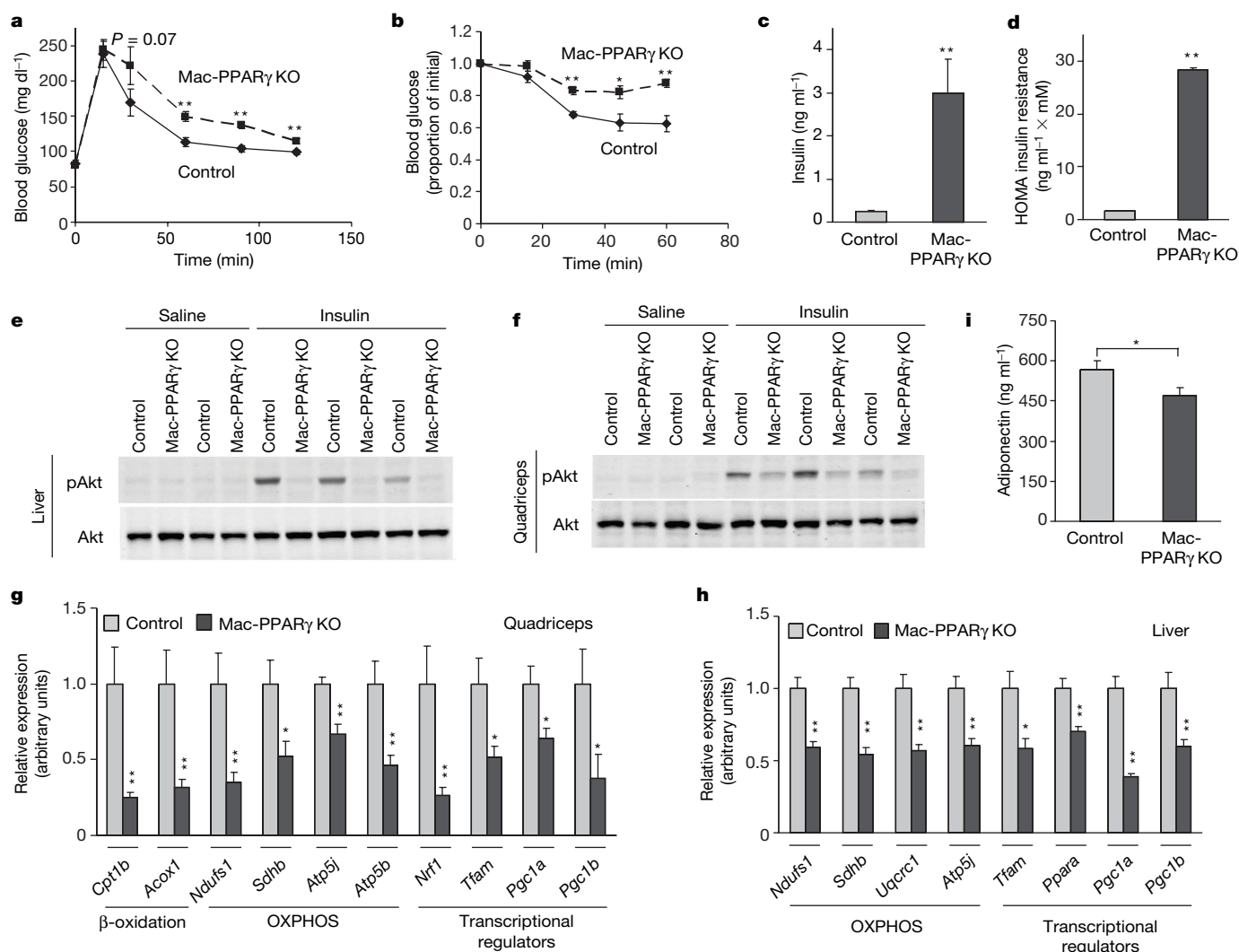


Figure 4 | Impaired glucose homeostasis in male Mac-PPAR γ KO mice fed a high fat diet. **a**, Oral glucose tolerance tests (1 g kg $^{-1}$) in male mice after 19 weeks of feeding the high fat diet ($n = 5$ per genotype). **b**, Insulin tolerance test. Obese mice were fasted for 4 h before intraperitoneal injection of insulin (0.65 U kg $^{-1}$). **c**, Fasting serum insulin levels in control and Mac-PPAR γ KO mice after a 4 h fast. **d**, The HOMA insulin resistance index of insulin sensitivity: insulin (ng ml $^{-1}$) \times glucose (mM). **e, f**, Decreased insulin

signalling in obese Mac-PPAR γ KO mice. Cellular lysates were immunoblotted for the total amount of the oncoprotein Akt and for serine phosphorylated (S473) Akt in liver (**e**) and quadriceps (**f**). **g, h**, Relative transcript levels of various genes involved in β -oxidation and oxidative phosphorylation, and of transcriptional regulators controlling these pathways in quadriceps (**g**) and liver (**h**). **i**, Circulating levels of adiponectin in control and Mac-PPAR γ KO mice. Data presented as mean \pm s.e.m. * $P < 0.05$; ** $P < 0.01$.

Nrf-1, *Pgc-1 α* and *Pgc-1 β* , was also reduced by 35–75% in the quadriceps of Mac-PPAR γ KO mice (Fig. 4g). Similarly, expression of genes in the electron transport chain and their transcriptional regulators was reduced by 30–60% in the livers of Mac-PPAR γ KO mice (Fig. 4h). Nonetheless, these livers did not display any overt signs of steatosis (Supplementary Fig. 6c, d), potentially reflecting their intact capacity to undergo β -oxidation (Supplementary Fig. 6e).

These findings suggest that direct or indirect effects of alternatively activated macrophages are important in maintaining oxidative capacity in skeletal muscle and liver. Because we did not observe significant differences in macrophage content or activation in liver and skeletal muscle of control and Mac-PPAR γ KO mice (Supplementary Fig. 6a, b and data not shown), we focused on factors secreted by adipocytes that modulate oxidative metabolism in peripheral tissues^{19,27}. Notably, recent studies have shown that adiponectin, a hormone specifically secreted by adipocytes, can induce PGC-1 α and mitochondrial biogenesis in skeletal muscle²⁸. Consistent with the global reduction in adipocyte function (Fig. 3c) and skeletal muscle oxidative capacity (Fig. 4g), the circulating level of adiponectin was reduced by ~18% in Mac-PPAR γ KO mice (Fig. 4i). In contrast, serum levels of resistin, total cholesterol and triglycerides were similar in both strains of mice (Supplementary Table 1).

In summary, the requirement for PPAR γ in expression of the alternatively activated phenotype, the absence of alternatively activated ATMs in Mac-PPAR γ KO mice, and the observation that Mac-PPAR γ KO mice are more susceptible to obesity and insulin resistance suggest that this program of macrophage activation protects against the metabolic consequences of obesity. In this setting of excess caloric intake, homeostatic functions performed by alternatively activated ATMs might allow animals to store and oxidize incoming lipids more efficiently, thereby maintaining insulin sensitivity and glucose tolerance. However, our findings raise additional issues that will require further investigation. First, to validate whether alternatively activated macrophages can be therapeutically exploited to treat type 2 diabetes, genetic and pharmacologic approaches that activate or inhibit IL-4 signalling in macrophages need to be tested. Second, although the presence of alternatively activated ATMs attenuates inflammation, we cannot exclude the possibility that remodelling activities of these cells might also lead to improvement in adipose tissue function. Lastly, in addition to their paracrine effects, it is plausible that peptides or lipids secreted by alternatively activated ATMs act in an endocrine fashion to modulate peripheral insulin sensitivity.

METHODS SUMMARY

For the generation of Mac-PPAR γ KO mice, we backcrossed the (floxed) PPAR $\gamma^{\text{fl/fl}}$ mice, generated by the Gonzalez laboratory²⁹, for ten generations onto the Th2-permissive Balb/c background. To generate mice in which the PPAR γ genes were disrupted in macrophages, PPAR $\gamma^{\text{fl/fl}}$ Balb/c mice were mated with LysM^{Cre} mice, also on the Balb/c background¹⁴. Cohorts of PPAR $\gamma^{\text{fl/fl}}$ (control) and PPAR $\gamma^{\text{fl/fl}}$; LysM^{Cre} (Mac-PPAR γ KO) mice were produced by intercrossing PPAR $\gamma^{\text{fl/fl}}$ and PPAR $\gamma^{\text{fl/fl}}$; LysM^{Cre} animals. Mice were genotyped for the presence of floxed and deleted alleles as described previously²⁹, and littermates were used to assemble the cohorts used in these studies.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 19 March; accepted 1 May 2007.

Published online 21 May 2007.

- Shoelson, S. E., Lee, J. & Goldfine, A. B. Inflammation and insulin resistance. *J. Clin. Invest.* **116**, 1793–1801 (2006).
- Hotamisligil, G. S. Inflammation and metabolic disorders. *Nature* **444**, 860–867 (2006).
- Weisberg, S. P. *et al.* Obesity is associated with macrophage accumulation in adipose tissue. *J. Clin. Invest.* **112**, 1796–1808 (2003).
- Xu, H. *et al.* Chronic inflammation in fat plays a crucial role in the development of obesity-related insulin resistance. *J. Clin. Invest.* **112**, 1821–1830 (2003).
- Kanda, H. *et al.* MCP-1 contributes to macrophage infiltration into adipose tissue, insulin resistance, and hepatic steatosis in obesity. *J. Clin. Invest.* **116**, 1494–1505 (2006).

- Kamei, N. *et al.* Overexpression of monocyte chemoattractant protein-1 in adipose tissues causes macrophage recruitment and insulin resistance. *J. Biol. Chem.* **281**, 26602–26614 (2006).
- Weisberg, S. P. *et al.* CCR2 modulates inflammatory and metabolic effects of high-fat feeding. *J. Clin. Invest.* **116**, 115–124 (2005).
- Gordon, S. & Taylor, P. R. Monocyte and macrophage heterogeneity. *Nature Rev. Immunol.* **5**, 953–964 (2005).
- Lumeng, C. N., Bodzin, J. L. & Saltiel, A. R. Obesity induces a phenotypic switch in adipose tissue macrophage polarization. *J. Clin. Invest.* **117**, 175–184 (2007).
- Gordon, S. Alternative activation of macrophages. *Nature Rev. Immunol.* **3**, 23–35 (2003).
- Lehrke, M. & Lazar, M. A. The many faces of PPAR γ . *Cell* **123**, 993–999 (2005).
- Huang, J. T. *et al.* Interleukin-4-dependent production of PPAR-gamma ligands in macrophages by 12/15-lipoxygenase. *Nature* **400**, 378–382 (1999).
- Vats, D. *et al.* Oxidative metabolism and PGC-1 β attenuate macrophage-mediated inflammation. *Cell Metab.* **4**, 13–24 (2006).
- Herbert, D. R. *et al.* Alternative macrophage activation is essential for survival during schistosomiasis and downmodulates T helper 1 responses and immunopathology. *Immunity* **20**, 623–635 (2004).
- Pauleau, A. L. *et al.* Enhancer-mediated control of macrophage-specific arginase I expression. *J. Immunol.* **172**, 7565–7573 (2004).
- Evans, R. M., Barish, G. D. & Wang, Y.-X. PPARs and the complex journey to obesity. *Nature Med.* **10**, 355–361 (2004).
- Alexander, J., Satoskar, A. R. & Russell, D. G. Leishmania species: models of intracellular parasitism. *J. Cell Sci.* **112**, 2993–3002 (1999).
- Holscher, C., Arendse, B., Schwegmann, A., Myburgh, E. & Brombacher, F. Impairment of alternative macrophage activation delays cutaneous leishmaniasis in nonhealing BALB/c mice. *J. Immunol.* **176**, 1115–1121 (2006).
- Rosen, E. D. & Spiegelman, B. M. Adipocytes as regulators of energy balance and glucose homeostasis. *Nature* **444**, 847–853 (2006).
- Rutschman, R. *et al.* Cutting edge: Stat6-dependent substrate depletion regulates nitric oxide production. *J. Immunol.* **166**, 2173–2177 (2001).
- Mootha, V. K. *et al.* PGC-1 α -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature Genet.* **34**, 267–273 (2003).
- Patti, M. E. *et al.* Coordinated reduction of genes of oxidative metabolism in humans with insulin resistance and diabetes: Potential role of PGC1 and NRF1. *Proc. Natl Acad. Sci. USA* **100**, 8466–8471 (2003).
- Petersen, K. F. *et al.* Mitochondrial dysfunction in the elderly: possible role in insulin resistance. *Science* **300**, 1140–1142 (2003).
- Vianna, C. R. *et al.* Hypomorphic mutation of PGC-1 β causes mitochondrial dysfunction and liver insulin resistance. *Cell Metab.* **4**, 453–464 (2006).
- Kelly, D. P. & Scarpulla, R. C. Transcriptional regulatory circuits controlling mitochondrial biogenesis and function. *Genes Dev.* **18**, 357–368 (2004).
- Lin, J., Handschin, C. & Spiegelman, B. M. Metabolic control through the PGC-1 family of transcription coactivators. *Cell Metab.* **1**, 361–370 (2005).
- Lazar, M. A. How obesity causes diabetes: not a tall tale. *Science* **307**, 373–375 (2005).
- Civitese, A. E. *et al.* Role of adiponectin in human skeletal muscle bioenergetics. *Cell Metab.* **4**, 75–87 (2006).
- Akiyama, T. E. *et al.* Conditional disruption of the peroxisome proliferator-activated receptor gamma gene in mice results in lowered expression of ABCA1, ABCG1, and apoE in macrophages and reduced cholesterol efflux. *Mol. Cell. Biol.* **22**, 2607–2619 (2002).
- Lumeng, C. N., Deyoung, S. M. & Saltiel, A. R. Macrophages block insulin action in adipocytes by altering expression of signaling and glucose transport proteins. *Am. J. Physiol. Endocrinol. Metab.* **292**, E166–E174 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank members of the Chawla laboratory for comments, and A. Loh and C. H. Lee for critiquing the manuscript. We also thank P. Murray, J. McKerrow and O. McGuinness for providing key reagents and technical guidance. This work was supported by grants made available to A.C. from the NIH, the Astellas Foundation, Takeda Pharmaceuticals North America, the Rockefeller Brothers Fund and by Goldman Philanthropic Partnerships; and to A.W.F. from the NIH and Columbia DERC. A.C. is a Charles E. Culpeper Medical Scholar. Support was provided by Stanford MSTP (to J.I.O. and A.R.E.), the AHA (to J.I.O.), a HHMI Gilliam fellowship (to A.R.E.), the NRSA (to R.R.R.-G.), and an NIH Training Grant (to L.M.). All animal care was in accordance with Stanford University's A-PLAC committee guidelines.

Author Contributions J.I.O. and R.R.R.-G. were involved in project planning, experimental work and data analysis; M.H.G., C.R.M., V.S., L.M., D.V. and A.R.E. performed experimental work; F.B. was involved in project planning; and A.W.F. and A.C. were involved in project planning, data analysis and manuscript preparation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.C. (achawla@stanford.edu).

METHODS

Functional analysis for alternative macrophage activation. Bone-marrow-derived macrophages from control, Mac-PPAR γ KO and PPAR $\delta^{-/-}$ (on the 129Sv background) mice were cultured as described in previously¹³. To promote alternative macrophage activation, bone-marrow-derived macrophages were stimulated with IL-4 (10 ng ml⁻¹) for 24–96 h. Arginase I activity was monitored via a colorimetric assay that detects the production of urea¹⁵. Fatty acid oxidation rates in macrophages were quantified using the sodium hydroxide trap in a modified tissue culture flask¹³. Cellular mitochondrial content was quantified using MitoTracker dyes (Molecular Probes), according to the manufacturer's instructions. All assays were performed in duplicate or triplicate and repeated at least three independent times, and total cell number or protein content was used for normalization of the data.

Gel mobility shift assays and transient transfections. *In vitro* translated proteins (TNT kit, Promega) and ³²P end-labelled oligonucleotides were used to carry out the gel mobility shift assays. Competition assays were performed with excess unlabelled oligonucleotides. The acyl-CoA oxidase PPAR response element served as a positive control for these studies. The sequence of the arginase I PPAR response element is (two half sites separated by a base pair): AAGTCA GAGAGCA. Transient transfection experiments were performed in RAW 264.7 cells. Briefly, cells were electroporated via the GenePulser II (BioRad) at 300 V, 1,000 μ F, and allowed to recover overnight. Luciferase assays were performed 16–20 h after stimulation with IL-4 (10 ng ml⁻¹) or rosiglitazone (1 μ M) with the Dual-Luciferase reporter assay system (Promega). The phRL-null plasmid (Promega) was used as an internal control to monitor transfection efficiency. All transfections were done in triplicate and repeated at least three times.

Immunoblots, ELISAs and immunohistochemistry. Total cellular proteins were immunoblotted for cytochrome C (1:1,000, BD Pharmingen), VDAC1 (1:2000, Molecular Probes), and β -actin (1:5,000, Sigma). Infrared dye (IR Dye 800) conjugated secondary antibodies (1:30,000, Rockland) were used for protein detection. For ELISAs, macrophages pre-treated with IL-4 (10 ng ml⁻¹) for 24 h were subsequently stimulated with lipopolysaccharide (5 ng ml⁻¹) for 6 h (TNF α) or 24 h (IL-6). Cytokine secretion was quantified by ELISA, according to the manufacturer's protocols (BD Pharmingen). All assays were done in triplicate and repeated at least three independent times. Macrophages in the gonadal fat pads were visualized by F4/80 immunostaining and quantified as described previously⁷.

Gene expression analysis. Trizol reagent (Invitrogen) was used to prepare total RNA from macrophages or tissues. Total RNA (2 μ g) was treated with DNase (1 U ml⁻¹) and reverse transcribed using a first-strand complementary DNA synthesis kit (Marligen). Q-PCR assays were carried out in triplicate on the DNA Engine Opticon 2 real-time PCR detection system. Relative expression levels for mRNAs were calculated using the comparative C_T method normalized to the L32 RNA.

Leishmania infection. *Leishmania major* (strain WHOM/IR/-/173) promastigotes were cultured at 27 °C in M199 containing 15% fetal bovine serum, 100 U ml⁻¹ penicillin, 2 mM glutamine, 100 μ M adenine, 5 μ g ml⁻¹ haemin, and 40 mM HEPES. Stationary phase metacyclic promastigotes (2×10^6) were injected into the left hind footpad¹⁸. Footpad thickness was measured weekly using a caliper (Mitutoyo), and reported as the difference in thickness between the infected and uninfected footpads. Footpad necrosis was graded using the following scale: 0, no visible abscesses; 1, one or more visible subcutaneous abscesses; 2, point surface necrosis; 3, surface necrosis over <50% of foot pad; 4, surface necrosis over >50% of foot pad; and 5, death of the foot.

Metabolic studies. Mice were fed a high fat diet (Bio-Serv F3282) to promote obesity. Oral glucose tolerance tests (1 g kg⁻¹) were performed after an overnight fast. For the insulin tolerance test, mice were fasted for 4 h before injection of human regular insulin (0.65 U kg⁻¹). Tail blood glucose levels were monitored at 0, 15, 30, 60 and 120 min time points with the Bayer Elite glucometer. For biochemical analysis of insulin signalling, obese mice of both genotypes were injected with insulin (5 mU g⁻¹) through their inferior vena cava after a 4 h fast. Liver and quadriceps were isolated 2 and 5 min, respectively, after insulin injection. Homogenized proteins were immunoblotted for total Akt and phospho-Akt (S473, Cell Signaling). Serum levels of adipokines (leptin, adiponectin and resistin), cytokines (TNF α , IL-6 and IL-1 β), lipids (total cholesterol and triglycerides), and insulin in fasted mice was quantified using commercially available kits.

Macrophage/adipocyte co-culture. Macrophage/adipocyte direct co-culture experiments were performed as described previously³⁰. Briefly, 100,000 bone-marrow-derived macrophages (day 10) were co-incubated with differentiated 3T3-L1 adipocytes (day 8) for 48 h. Cultures were stimulated with insulin (100 nM) for 30 min, and glucose uptake was monitored using 0.2 mM 2-[³H]

deoxyglucose (1 μ Ci ml⁻¹) over 5 min. Incorporated radioactivity was normalized for protein content, and expressed as fold increase over untreated samples. **Statistical analysis.** Data are presented as mean \pm s.e.m., and Student's *t*-test (two-tailed distribution, two-sample unequal variance) was used to calculate the *P* value. Statistical significance is displayed as *P* < 0.05 (one asterisk) or *P* < 0.01 (two asterisks).

Ambra1 regulates autophagy and development of the nervous system

Gian Maria Fimia^{1*}, Anastassia Stoykova^{2*}, Alessandra Romagnoli^{1*}, Luigi Giunta^{3,5}, Sabrina Di Bartolomeo^{3,5}, Roberta Nardacci¹, Marco Corazzari¹, Claudia Fuoco^{3,5}, Ahmet Ucar², Peter Schwartz⁶, Peter Gruss², Mauro Piacentini^{1,4}, Kamal Chowdhury² & Francesco Cecconi^{3,5}

Autophagy is a self-degradative process involved both in basal turnover of cellular components and in response to nutrient starvation or organelle damage in a wide range of eukaryotes^{1–3}. During autophagy, portions of the cytoplasm are sequestered by double-membraned vesicles called autophagosomes, and are degraded after fusion with lysosomes for subsequent recycling⁴. In vertebrates, this process acts as a pro-survival or pro-death mechanism in different physiological and pathological conditions, such as neurodegeneration and cancer^{2,5–7}; however, the roles of autophagy during embryonic development are still largely uncharacterized³. Beclin1 (Becn1; coiled-coil, myosin-like BCL2-interacting protein) is a principal regulator in autophagosome formation, and its deficiency results in early embryonic lethality^{8,9}. Here we show that Ambra1 (activating molecule in Beclin1-regulated autophagy), a large, previously unknown protein bearing a WD40 domain at its amino terminus, regulates autophagy and has a crucial role in embryogenesis. We found that Ambra1 is a positive regulator of the Becn1-dependent programme of autophagy, as revealed by its overexpression and by RNA interference experiments *in vitro*. Notably, Ambra1 functional deficiency in mouse embryos leads to severe neural tube defects associated with autophagy impairment, accumulation of ubiquitinated proteins, unbalanced cell proliferation and excessive apoptotic cell death. In addition to identifying a new and essential element regulating the autophagy programme, our results provide *in vivo* evidence supporting the existence of a complex interplay between autophagy, cell growth and cell death required for neural development in mammals.

Autophagy is a highly regulated mechanism, as demonstrated by the identification of several autophagy-related (Atg) genes in the yeast *Saccharomyces cerevisiae*¹⁰. Recent studies have not only demonstrated the evolutionary conservation of Atg function in vertebrates but also highlighted the involvement of the autophagic machinery in many aspects of tissue homeostasis^{2,3}. Indeed, autophagy has a protective function against the onset of neurodegeneration^{5,6} and the alteration of cell growth in cancer^{7,9,11}. Moreover, a complex interaction between the pathways of autophagy and apoptosis has been reported in several *in vitro* systems, in which dysregulation of the apoptotic pathway induces changes to the autophagy programme and vice versa^{12–15}.

A role of autophagy in the regulation of vertebrate development has long been proposed by studies reporting morphological features of autophagy during embryogenesis^{3,16–18}. Targeted mutagenesis of a few Atg genes in mice substantiated this hypothesis. Embryos that

contain a mutation to Becn1, the orthologue of yeast Atg6, die early during embryogenesis, exhibiting a developmental delay⁹, whereas Atg5- or Atg7-deficient mice survive embryogenesis but suffer from nutrient and energy insufficiency soon after birth^{19,20}. In addition, Atg5^{-/-} embryos were shown to have deficient autophagy-dependent clearance of apoptotic cell corpses during development²¹. However, it is not yet known how autophagy can be regulated during development, and whether vertebrate-specific factors besides the Atg genes have a role in this context.

Using a large-scale mutagenesis approach based on gene trapping in mice²², we isolated a novel gene involved in the development of the nervous system. This gene encodes a 1,300-amino-acid-long WD40 protein that is highly conserved in vertebrates (Supplementary Fig. 1). We named this protein Ambra1 because of its functions, as described below. In embryos homozygous for the gene-trapped allele (*Ambra1*^{gt/gt}) we revealed a fusion transcript (trapped gene-*lacZ*) and its corresponding protein, whereas the endogenous protein was absent (Fig. 1a, b). Because β -galactosidase (β -gal) activity in a gene-trap line mimics the expression of the tagged endogenous gene²³, we performed *Ambra1* expression analysis after whole-mount β -gal staining in heterozygous embryos throughout development and in postnatal brain (see Fig. 1c–e and Supplementary Fig. 2). At embryonic day (E)8.5, strong staining was detected throughout the neuroepithelium (Fig. 1c and Supplementary Fig. 2a). At E11.5 (Fig. 1d, e), a robust expression was seen in the ventral-most part of the spinal cord, the encephalic vesicles, the neural retina, the limbs and the dorsal root ganglia. Messenger RNA *in situ* hybridization using an *Ambra1* probe confirmed the expression pattern (Fig. 1f and Supplementary Fig. 2b, compared with β -gal staining in Supplementary Fig. 2c). At later developmental stages, the β -gal staining became abundant in the entire developing nervous system as well as in other tissues (Supplementary Fig. 2d); strong expression was observed in the cortex, hippocampus and striatum of postnatal brain (Supplementary Fig. 2e–i).

Notably, we found that homozygosity of the *Ambra1* mutation causes embryonic lethality in outbred or inbred genetic backgrounds (Supplementary Table 1). Most *Ambra1*^{gt/gt} embryos at stages E10–E14.5 exhibit neural tube defects detected as midbrain/hindbrain exencephaly and/or spina bifida (Fig. 1g–j). Histological analysis revealed that, in the embryonic forebrain of *Ambra1*^{gt/gt} mice, the vesicles are closed but displaced, the proliferative neuroepithelium shows an extensive overgrowth and the spinal cord is enlarged (Fig. 1k, l). Next, we performed a detailed analysis of cell death, differentiation and growth in *Ambra1*^{gt/gt} embryos. Excessive

¹National Institute for Infectious Diseases IRCCS ‘L. Spallanzani’, 00149 Rome, Italy. ²Department of Molecular Cell Biology, Max Planck Institute of Biophysical Chemistry, 37077 Göttingen, Germany. ³Dulbecco Telethon Institute at the Department of Biology, ⁴Department of Biology, University of Rome ‘Tor Vergata’, 00133 Rome, Italy. ⁵Laboratory of Molecular Neuroembryology, IRCCS Fondazione Santa Lucia, 00143 Rome, Italy. ⁶Department of Anatomy and Embryology, University of Göttingen, 37085 Göttingen, Germany.

*These authors contributed equally to this work.

apoptosis was present in selected areas of the mutant nervous system from E9.0 onwards (Fig. 1m–p, Supplementary Fig. 3 and data not shown). We then observed that patterning regulators, such as Shh, Fgf8, En1, En2, Gli3, Pax2, Pax3, Pax5, Pax6, Ng2 (also called Neurog2), Dlx1, Dlx2 and Mash1 (Ascl1), are present at wild-type levels in the developing mutant brain and spinal cord, although a few of them are displaced or show decreased expression (Supplementary Fig. 4 and data not shown). Finally, 5-bromodeoxyuridine (BrdU)-uptake experiments and measurement of mitoses revealed a statistically significant increase of proliferating cells in *Ambra1*^{gt/gt} E8.5 neural folds (Fig. 1q, r and Supplementary Fig. 5a, b). However, from E9.5 onwards, the percentage of BrdU-positive cells out of the total number of cells was similar between the mutant and wild-type neuroepithelium, indicating that the hyperproliferation phenotype appears at the onset of neurulation (Fig. 1s, t and Supplementary Fig. 5a). Taken together, these results indicate that *Ambra1* is necessary to control cell proliferation and to guarantee cell survival during nervous system development.

To define the biological processes regulated by *Ambra1*, we identified its molecular interactors using a yeast two-hybrid approach. Complementary DNA encoding the first 667 amino acids of the

human *AMBRA1* orthologue (Supplementary Fig. 1c) was used to screen a human brain cDNA expression library. One of the isolated clones (Fig. 2a) encodes the Atg protein BECN1, a component of the class III phosphatidylinositol-3-OH kinase (also known as Vps34) complex regulating autophagosome formation in mammals^{8,24}. The BECN1–*AMBRA1* interaction was confirmed in co-transduced human 2FTGH fibroblasts by co-immunoprecipitation assays (Fig. 2b, c). To map the region responsible for BECN1 binding, we tested different *AMBRA1* mutant constructs in co-immunoprecipitation with BECN1. A central region of the protein (F2) is necessary and sufficient for an effective interaction with BECN1, whereas the *AMBRA1* amino terminus (F1) shows a minimal binding capability, and its carboxy terminus (F3) does not interact at all (Fig. 2b). The BECN1-associated kinase Vps34 co-immunoprecipitates with *AMBRA1*, suggesting that BECN1, Vps34 and *AMBRA1* are components of a multiprotein complex (Fig. 2c). The BECN1–*AMBRA1* interaction was also confirmed in developing brain samples by co-immunoprecipitating the endogenous proteins (Fig. 2d). Consistent with their interaction, BECN1 and *AMBRA1* showed a vesicular-like staining that mostly co-localized in 2FTGH cells (Fig. 2e).

The BECN1–*AMBRA1* interaction prompted us to test for a possible role of *AMBRA1* in autophagy regulation. The occurrence of autophagy was analysed in 2FTGH cells by assessing both conversion of microtubule-associated protein 1 light-chain 3 (LC3)-I to LC3-II and its translocation to autophagic structures—two sequential steps in autophagosome formation²⁵. First we tested the cell response to two autophagic stimuli—the mTOR (mammalian target of rapamycin)-inhibitor rapamycin and nutrient deprivation—upon reduction of *AMBRA1* levels by RNA interference. Downregulation of *AMBRA1* (Fig. 3a) resulted in a remarkable decrease in autophagy (Fig. 3b, c and Supplementary Fig. 6a). Because BECN1 overexpression has been shown to induce autophagy *per se*⁸, we tested the effect of *AMBRA1* overexpression on autophagy either at basal levels or in rapamycin-treated 2FTGH cells. A significant increase of both basal and rapamycin-induced autophagy was observed (Fig. 3d). Furthermore, we checked whether interaction with BECN1 was required for *AMBRA1*-induced autophagy by using *AMBRA1* F1, F2 and F3 constructs (see Fig. 2b). F2 was able to induce autophagy

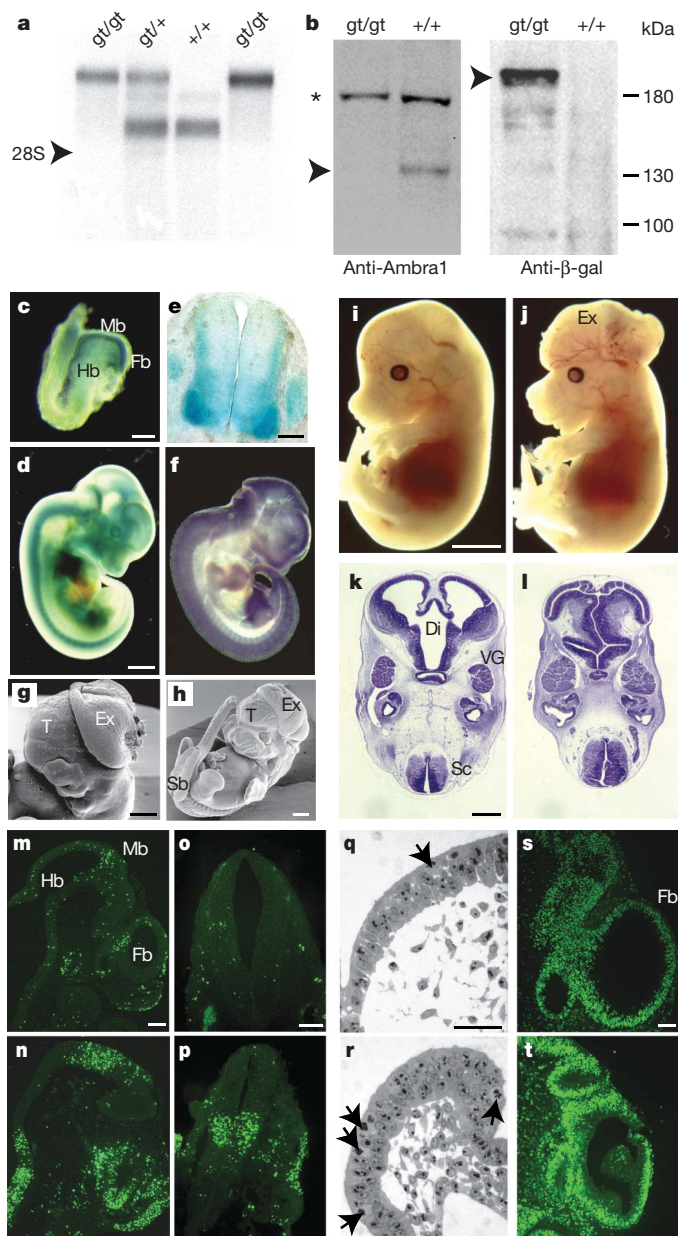


Figure 1 | Neural tube defects in *Ambra1* mutant embryos. **a**, Detection of wild-type (+) and *Ambra1* gene-trap transcripts (gt) from total E14.5 embryos analysed by northern blotting. Size marker, 28S ribosomal RNA. **b**, Immunoblot analysis of E14.5 embryo brain extracts using antibodies against Ambra1 (left panel) and β -gal (right panel). Specific bands (arrowheads) and unspecific signals (asterisk) are indicated. Molecular weights are measured in kilodaltons (kDa). The fusion protein can be revealed only by means of the more sensitive anti- β -gal antibody. **c–f**, Expression of *Ambra1* in the mouse embryonic nervous system. β -gal-staining on whole-mount *Ambra1*^{+/gt} mouse embryos at E8.5 (**c**), E11.5 (**d**) and on a cross-section of E11.5 spinal cord (**e**) is shown. Fb, forebrain; Hb, hindbrain; Mb, midbrain. **f**, Matching expression of the endogenous gene was revealed by whole-mount *Ambra1* mRNA *in situ* hybridization analysis in E11.5 wild-type embryos. **g–l**, Electron scanning microscopy analysis of E11.5 (**g**) and E12.5 (**h**) *Ambra1*^{gt/gt} embryos. Note the failure of the neural tube closure, the extensive midbrain/hindbrain exencephaly (Ex) with a closed telencephalon (T), and the lumbosacral spina bifida (Sb). **i, j**, E14.5 wild-type (**i**) and *Ambra1*^{gt/gt} (**j**) embryos are characterized by prominent exencephaly. **k, l**, Histological analysis of E12.5 wild-type (**k**) and *Ambra1*^{gt/gt} (**l**) embryos on cross-section. Note the absence of a normal ventricular system, the extensive overgrowth of the proliferative neuroepithelium in the diencephalon (Di) and spinal cord (Sc), and the enlarged fifth ganglia (VG) in the *Ambra1*^{gt/gt} embryo. **m–p**, TUNEL (TdT-mediated dUTP nick end labelling) staining of E10.5 brain (**m, n**) and E9.5 spinal cord (**o, p**) in wild-type (**m, o**) and *Ambra1*^{gt/gt} (**n, p**) embryos. **q–t**, Analysis of cell proliferation in wild-type (**q, s**) and *Ambra1*^{gt/gt} (**r, t**) embryos on transverse sections of E8.5 cephalic neural folds in prospective hindbrain region (mitoses, arrows in **q, r**) and on sagittal sections of E10.5 forebrain (BrdU uptake, **s, t**). Scale bars: **c, g, h, k, l**, 500 μ m; **d**, 1 mm; **e**, 150 μ m; **i, j**, 2 mm; **m, s**, 400 μ m; **o**, 78 μ m; and **q**, 50 μ m.

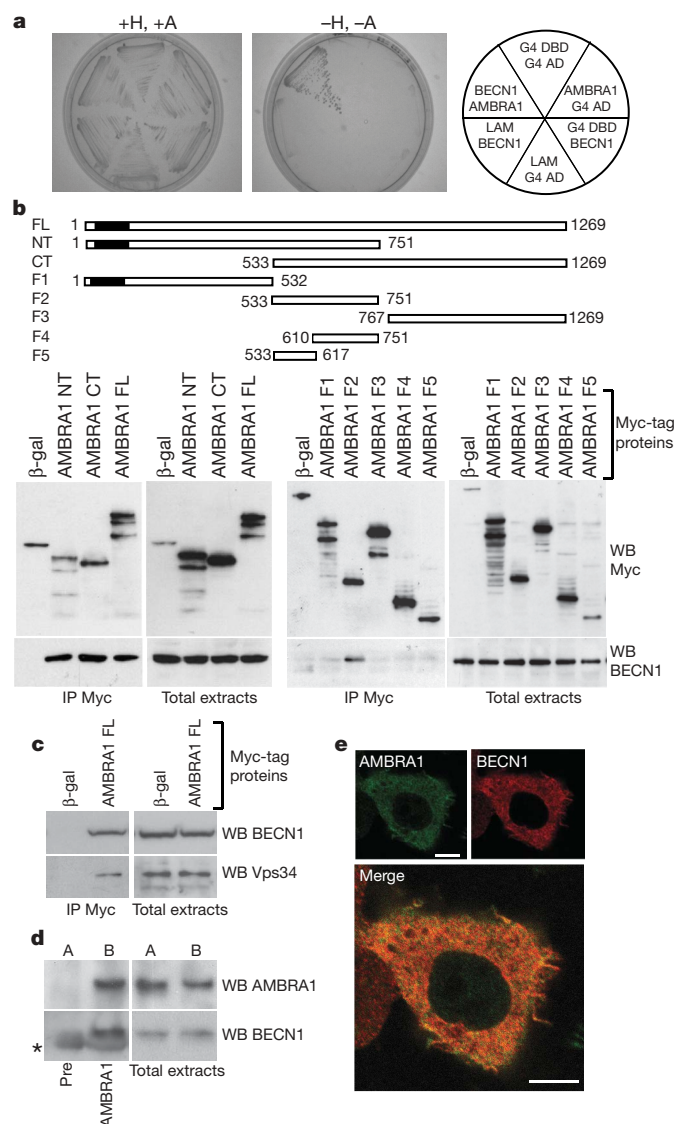


Figure 2 | AMBRA1 is a novel BECN1-interacting protein.

a, AMBRA1–BECN1 interaction by yeast two-hybrid assay. Yeast cells were co-transfected with the indicated plasmids and plated in medium with (+) or without (–) histidine (H) and adenine (A). AD, activation domain; DBD, DNA-binding domain; G4, Gal4; LAM, lamin. **b**, AMBRA1–BECN1 interaction in mammalian cells. 2FTGH cells were co-infected with retroviral vectors encoding BECN1 and the indicated Myc-tagged AMBRA1 proteins, or Myc-tagged β-gal as a negative control. CT, C terminal; FL, full length; NT, N terminal. Protein extracts were immunoprecipitated using an anti-Myc-tagged antibody (IP Myc). Purified complexes and corresponding total extracts were analysed by western blot using anti-Myc (WB Myc, upper panels) or anti-BECN1 (WB BECN1, lower panels) antibodies. When multiple bands are present, the upper band has the expected molecular mass. A scheme of the various AMBRA1 mutants is reported. **c**, AMBRA1 is in a complex that includes BECN1 and Vps34. Protein extracts from 2FTGH cells co-expressing BECN1 and the Myc-tagged AMBRA1 full-length protein (FL) or Myc-tagged β-gal were immunoprecipitated using an anti-Myc-tag antibody (IP Myc). Purified complexes and corresponding total extracts were analysed using anti-BECN1 (WB BECN1) and anti-Vps34 (WB Vps34) antibodies. **d**, AMBRA1–BECN1 interaction in developing brain tissues. Protein extracts from two E14.5 wild-type brains (A and B) were immunoprecipitated using an anti-AMBRA1 antibody (AMBRA1) or the pre-immune serum (Pre). Purified complexes and corresponding total extracts were analysed using anti-AMBRA1 (WB AMBRA1) and anti-BECN1 (WB BECN1) antibodies. The asterisk indicates the position of co-migrating immunoglobulins. **e**, BECN1 and AMBRA1 co-localize in mammalian cells. 2FTGH cells co-expressing BECN1 and Myc-tagged AMBRA1 were stained for Myc (green) and BECN1 (red). Scale bars, 8 μm.

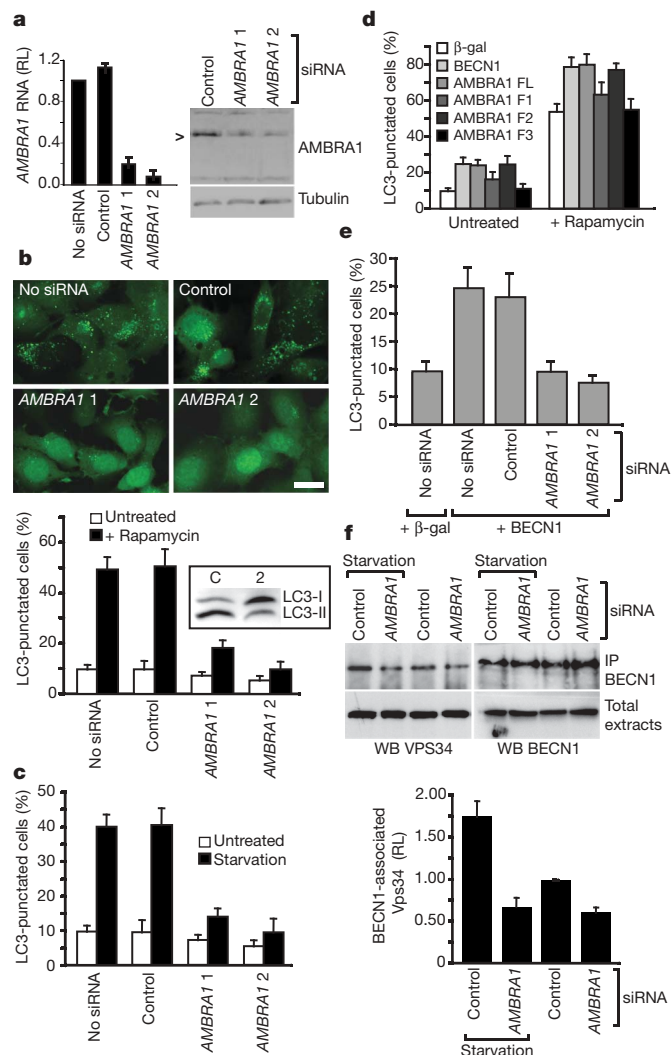


Figure 3 | AMBRA1 regulates autophagy. **a**, AMBRA1 downregulation in 2FTGH cells using specific short interfering RNA (siRNA) oligonucleotides (1 and 2). AMBRA1 mRNA and protein levels were analysed by quantitative polymerase chain reaction (PCR, left) and western blotting (right), respectively. An unrelated oligonucleotide was used as the control. Tubulin was used as the protein-loading control. RL, relative levels. The arrowhead indicates the AMBRA1-specific band. **b**, Rapamycin-induced autophagy requires AMBRA1. After AMBRA1 downregulation, 2FTGH cells were treated with rapamycin; after 48 h, the occurrence of autophagy was analysed by appearance of GFP–LC3 punctate staining or by LC3-I to LC3-II conversion (western blot in inset). Representative results are accompanied by a graph reporting data from three experiments. Scale bar, 20 μm. **c**, Nutrient-starvation-induced autophagy requires AMBRA1. After AMBRA1 downregulation, 2FTGH cells were starved for 4 h and analysed for appearance of GFP–LC3 punctate staining. **d**, AMBRA1 overexpression increases basal and rapamycin-induced autophagy. 2FTGH cells were transduced with full-length AMBRA1 (FL), AMBRA1 fragments (F1–3), BECN1 and β-gal (negative control) encoding retroviruses. They were then stimulated with rapamycin or left untreated, and analysed by appearance of GFP–LC3 punctate staining. **e**, BECN1-induced autophagy requires AMBRA1. After AMBRA1 downregulation, GFP–LC3-expressing 2FTGH cells were transduced with a BECN1-encoding retrovirus. Occurrence of autophagy was analysed 48 h later by appearance of GFP–LC3 punctate staining. **f**, AMBRA1 downregulation reduces the amount of Vps34 associated to BECN1 during autophagy. Forty-eight hours after AMBRA1 downregulation with oligo 2, 2FTGH cells were either starved or cultured in standard medium for an additional 4 h. Protein extracts were immunoprecipitated (IP) using an anti-BECN1 antibody. Purified complexes and corresponding total extracts were analysed by western blotting with anti-Vps34 (left panels) and anti-BECN1 (right panels) antibodies. The graph indicates the signal intensity of BECN1-associated Vps34 as determined by densitometry. Values in **a–f** represent the mean ± s.d. of three experiments.

at levels similar to the full-length AMBRA1, F3 had no detectable effect (Fig. 3d), whereas F1 showed an intermediate effect, probably because of its residual BECN1-binding capability. The effects of AMBRA1 dysregulation on the autophagy process were confirmed by measurement of the formation of acidic vesicular organelles²⁶ (Supplementary Fig. 6b, c). Also we observed that BECN1-mediated induction of autophagy was markedly reduced after AMBRA1 down-regulation (Fig. 3e). Consistently, we showed that AMBRA1 down-regulation leads to a reduced capability of BECN1 to interact with its associated kinase Vps34, and to a decrease in Vps34 activity in cells in which autophagy has been induced (Fig. 3f and Supplementary Fig. 6d–f). This finding suggests a role for AMBRA1 in favouring the BECN1–Vps34 functional interaction. Taken together, these results show that AMBRA1 is a key factor in autophagy regulation and is required for BECN1 activity.

The autophagy-promoting activity of BECN1 has been associated with inhibition of cell proliferation⁸. Given the observed hyperproliferative phenotype in *Ambra1* mutant embryos (see Fig. 1q–t and Supplementary Fig. 5), we investigated whether AMBRA1 dosage is critical for cell growth control *in vitro*. AMBRA1 downregulation or its overexpression resulted in a significant increase or decrease of cell proliferation rate, respectively (Supplementary Fig. 7a, b). Moreover, inhibition of proliferation by AMBRA1 is dependent on BECN1, as this effect is abolished when BECN1 is downregulated (Supplementary Fig. 7c). These observations support the view that dysregulation of BECN1-dependent autophagy is linked to abnormal cell proliferation.

On the basis of these *in vitro* results, we investigated whether autophagy defects were detectable in *Ambra1* mutants. Conversion of LC3-I to LC3-II was markedly reduced in *Ambra1*^{gt/gt} E14.5 embryos (Fig. 4a, left panel). LC3 translocation was analysed by crossing *Ambra1*^{+/gt} mice with transgenic mice ubiquitously expressing green fluorescent protein (GFP)–LC3 (ref. 27). Fluorescence microscopy examination of E10.5 neuroepithelium showed clusters of GFP–LC3 dots in wild-type embryos, whereas a diffuse signal was detectable in *Ambra1*^{gt/gt} littermates (Fig. 4a, right panels). Next, because autophagy impairment in the adult nervous system is accompanied by accumulation of ubiquitinated proteins^{5,6}, we checked for ubiquitin expression in developing brains of *Ambra1*^{gt/gt} embryos. Numerous ubiquitin-positive cells were present in the mutant neuroepithelium, with a predominant nuclear staining (Fig. 4b). Furthermore, autophagy defects were confirmed in *Ambra1*^{gt/gt} mouse embryonic fibroblasts, as revealed by GFP–LC3 translocation to autophagosomal membranes, ultrastructural analysis and measurement of formation of acidic vesicular organelles (Fig. 4c, d, Supplementary Fig. 8a–c). Finally, a reduction in the activity of the mutant *Ambra1* protein present in the *Ambra1*^{gt/gt} mice, in terms of the capability to both interact with Becln1 and induce autophagy, was confirmed by transfection experiments using a construct mimicking the mutant fusion mRNA (Supplementary Fig. 8d, e). Thus, we conclude that the *Ambra1* gene-trap allele carries a loss-of-function mutation that causes a severe autophagy defect in *Ambra1*^{gt/gt} embryos.

Here we show that deficiency of *Ambra1*, a novel *Becln1* partner mainly expressed in the neural tube during development, impairs autophagy and results in severe neural tube defects. Besides the putative role of AMBRA1 in human congenital brain malformation, its expression profile in adult brain compartments, which are affected in patients with neurodegenerative disorders (Supplementary Fig. 2), suggests a role for AMBRA1 in these diseases, as already shown for other autophagy regulators^{5,6}.

Morphogenesis of the nervous system is based on a complex interplay between the control of balanced cell proliferation, differentiation and apoptosis²⁸. Our finding that the level of ubiquitinated proteins is markedly enhanced in *Ambra1* mutant neuroepithelium implies a role for autophagy-dependent protein turnover in the control of neural development. A sensitive balance of degradation of key neurodevelopmental regulators might be hypothesized as having a

role in the determination of cell fate and, when altered, may lead *per se* to the observed neural tube defects. Alternatively, such a phenotype might be the consequence of disturbed cell proliferation and apoptosis, caused by autophagy dysregulation. In fact, *Ambra1* deficiency during embryogenesis leads to excessive cell proliferation at early stages followed by increased apoptosis in the neuroepithelium. Genetic evidence that autophagy controls cell proliferation was provided by experiments investigating dysregulation of *Becln1* and its

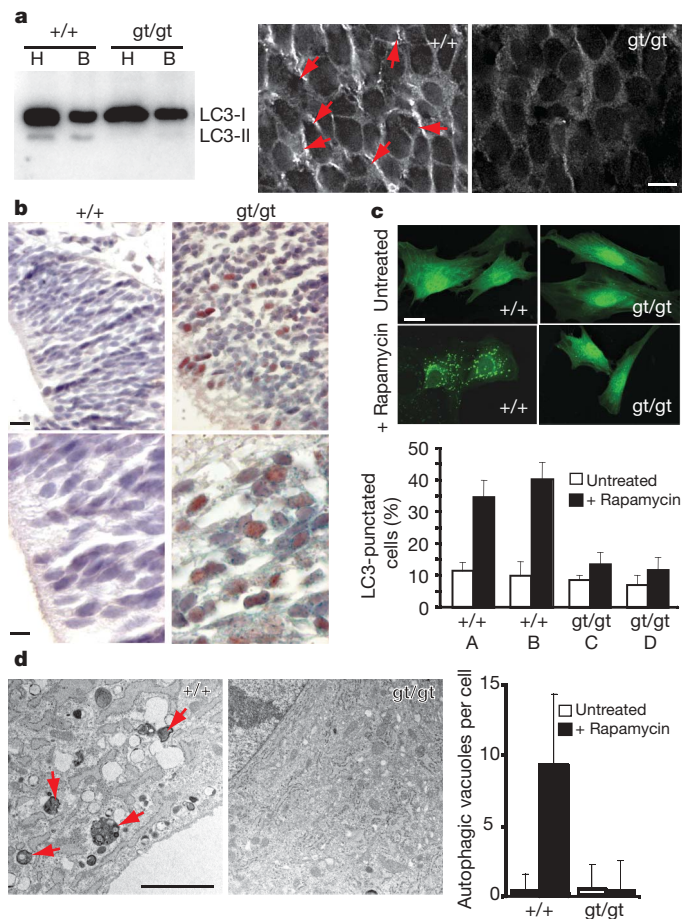


Figure 4 | Autophagy is impaired in *Ambra1* mutant embryos. **a**, Alteration of both LC3 conversion and translocation in *Ambra1*^{gt/gt} embryos. Left panel: immunoblot analysis of LC3 in E14.5 embryos. LC3-I to LC3-II conversion is reduced in mutant extracts (LC3-II/LC3-I ratio, as measured by densitometry, from left to right: 5%, 9.8%, 2.5%, 4.6%). B, body; H, head. Centre and right panels: sections from E10.5 *Ambra1*^{+/gt};GFP–LC3 and *Ambra1*^{gt/gt};GFP–LC3 neuroepithelia were compared to evaluate GFP–LC3 subcellular distribution. Clusters of GFP–LC3 punctate structures (arrows) are indicated in wild-type sections. Images were optimized by deconvolution. Scale bar, 10 μm. **b**, Ubiquitin-positive cells in developing brains of *Ambra1*^{gt/gt} embryos. Sagittal sections from E11.5 embryos were stained using an anti-ubiquitin antibody. Ubiquitin accumulates in numerous mutant neuroepithelial cells (shown at different magnification: upper panels, scale bar, 25 μm; lower panels, scale bar, 10 μm). *Ambra1* mutant protein is visualized in the left panels by β-gal activity (light-blue staining). Counterstaining, haematoxylin blue. **c**, Rapamycin-induced autophagy is impaired in *Ambra1*^{gt/gt} murine embryonic fibroblasts (MEFs). Cells dissected from embryos with different genotypes (A–D) were treated with rapamycin, and the occurrence of autophagy was analysed 48 h later by the appearance of GFP–LC3 punctate staining. Representative results are shown in the upper panels, whereas the lower panel shows the data from three experiments. Scale bar, 20 μm. **d**, Ultrastructural analysis by means of electron microscopy of ultrathin sections from wild-type and *Ambra1*^{gt/gt} cultured MEFs. The arrows indicate autophagic vacuoles. Scale bar, 1.5 μm. Quantification of the mean number of autophagic vacuoles per cell in wild-type and *Ambra1*^{gt/gt} cultured MEFs is shown in the graph. Values in **c** and **d** represent the mean ± s.d. of three experiments.

interactor ultraviolet radiation resistance associated gene (UVRAG)^{8,9,11,29}. Our results suggest that the excessive cell proliferation in the *Ambra1* mutant is directly associated with *Becn1* dysregulation. However, we cannot rule out that neuroepithelial cells deficient in autophagy undergo apoptosis in a cell-autonomous fashion, whereas the surrounding cells, lacking proper signalling (for example, Shh-mediated mitogenic induction), overproliferate.

The identification of a vertebrate-specific autophagy regulator that is active in an evolutionarily conserved pathway opens up a new scenario involving autophagy in specialized roles during the development of higher eukaryotes.

METHODS SUMMARY

Generation of *Ambra1*^{gst/gst} mice. The embryonic stem cell line containing an insertion of the gene-trap vector pGT1.8geo (a gift from W. Skarnes)²³ within the *Ambra1* gene was generated as described²². Founder chimaeric males were mated with outbred NMRI females, and heterozygous progenies were mated to maintain the allele. Also, backcrossing with C57BL/6 males was carried out to obtain an inbred strain. In some experiments, GFP-LC3 mice²⁷ were crossed with *Ambra1*^{+/gst} mice to generate *Ambra1*^{+/gst};GFP-LC3 mice. The latter were then crossed with each other to obtain *Ambra1*^{gst/gst};GFP-LC3 embryos.

Autophagy assays. Autophagy was measured in GFP-LC3 transduced cells as described³⁰. Briefly, cells grown on coverslips were fixed with 4% paraformaldehyde in PBS, washed three times and directly examined by confocal microscopy. The results indicate the percentage of GFP-LC3-positive cells with GFP-LC3 punctate dots. A minimum of 50–100 cells per sample were counted for triplicate samples per condition per experiment.

To quantify the development of acidic vesicular organelles²⁶, cells were detached by trypsin digestion, washed with PBS, stained with acridine orange (1 µg ml⁻¹; Sigma-Aldrich) for 15 min and analysed using a FACScan flow cytometer (Becton Dickinson) and CellQuest software. For electron microscopy, cells were fixed with 2.5% glutaraldehyde in 0.1 M cacodylate buffer pH 7.4 for 45 min at 4 °C, rinsed in cacodylate buffer, postfixed in 1% OsO₄ in cacodylate buffer, dehydrated and embedded in Epon. Ultrathin sections were briefly contrasted with uranyl acetate and photographed with a Zeiss EM900 electron microscope.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 30 January; accepted 2 May 2007.

Published online 24 June 2007.

- Lum, J. J., DeBerardinis, R. J. & Thompson, C. B. Autophagy in metazoans: cell survival in the land of plenty. *Nature Rev. Mol. Cell Biol.* **6**, 439–448 (2005).
- Levine, B. & Yuan, J. Autophagy in cell death: an innocent convict? *J. Clin. Invest.* **115**, 2679–2688 (2005).
- Levine, B. & Klionsky, D. J. Development by self-digestion: molecular mechanisms and biological functions of autophagy. *Dev. Cell* **6**, 463–477 (2004).
- Klionsky, D. J. The molecular machinery of autophagy: unanswered questions. *J. Cell Sci.* **118**, 7–18 (2005).
- Hara, T. *et al.* Suppression of basal autophagy in neural cells causes neurodegenerative disease in mice. *Nature* **441**, 885–889 (2006).
- Komatsu, M. *et al.* Loss of autophagy in the central nervous system causes neurodegeneration in mice. *Nature* **441**, 880–884 (2006).
- Edinger, A. L. & Thompson, C. B. Defective autophagy leads to cancer. *Cancer Cell* **4**, 422–424 (2003).
- Liang, X. H. *et al.* Induction of autophagy and inhibition of tumorigenesis by beclin 1. *Nature* **402**, 672–676 (1999).
- Yue, Z., Jin, S., Yang, C., Levine, A. J. & Heintz, N. Beclin 1, an autophagy gene essential for early embryonic development, is a haploinsufficient tumor suppressor. *Proc. Natl Acad. Sci. USA* **100**, 15077–15082 (2003).
- Huang, W. P. & Klionsky, D. J. Autophagy in yeast: a review of the molecular machinery. *Cell Struct. Funct.* **27**, 409–420 (2002).
- Qu, X. *et al.* Promotion of tumorigenesis by heterozygous disruption of the beclin 1 autophagy gene. *J. Clin. Invest.* **112**, 1809–1820 (2003).

- Shimizu, S. *et al.* Role of Bcl-2 family proteins in a non-apoptotic programmed cell death dependent on autophagy genes. *Nature Cell Biol.* **6**, 1221–1228 (2004).
- Boya, P. *et al.* Inhibition of macroautophagy triggers apoptosis. *Mol. Cell. Biol.* **25**, 1025–1040 (2005).
- Lum, J. J. *et al.* Growth factor regulation of autophagy and cell survival in the absence of apoptosis. *Cell* **120**, 237–248 (2005).
- Yousefi, S. *et al.* Calpain-mediated cleavage of Atg5 switches autophagy to apoptosis. *Nature Cell Biol.* **8**, 1124–1132 (2006).
- Clarke, P. G. Developmental cell death: morphological diversity and multiple mechanisms. *Anat. Embryol.* **181**, 195–213 (1990).
- Baehrecke, E. H. How death shapes life during development. *Nature Rev. Mol. Cell Biol.* **3**, 779–787 (2002).
- Lockshin, R. A. & Zakeri, Z. Apoptosis, autophagy, and more. *Int. J. Biochem. Cell Biol.* **36**, 2405–2419 (2004).
- Kuma, A. *et al.* The role of autophagy during the early neonatal starvation period. *Nature* **432**, 1032–1036 (2004).
- Komatsu, M. *et al.* Impairment of starvation-induced and constitutive autophagy in Atg7-deficient mice. *J. Cell Biol.* **169**, 425–434 (2005).
- Qu, X. *et al.* Autophagy gene-dependent clearance of apoptotic cells during embryonic development. *Cell* **128**, 931–946 (2007).
- Stoykova, A., Chowdhury, K., Bonaldo, P., Torres, M. & Gruss, P. Gene trap expression and mutational analysis for genes involved in the development of the mammalian nervous system. *Dev. Dyn.* **212**, 198–213 (1998).
- Skarnes, W. C., Moss, J. E., Hurlley, S. M. & Beddington, R. S. Capturing genes encoding membrane and secreted proteins important for mouse development. *Proc. Natl Acad. Sci. USA* **92**, 6592–6596 (1995).
- Kihara, A., Kabeya, Y., Ohsumi, Y. & Yoshimori, T. Beclin-phosphatidylinositol 3-kinase complex functions at the trans-Golgi network. *EMBO Rep.* **2**, 330–335 (2001).
- Kabeya, Y. *et al.* LC3, a mammalian homologue of yeast Apg8p, is localized in autophagosome membranes after processing. *EMBO J.* **19**, 5720–5728 (2000).
- Paglin, S. *et al.* A novel response of cancer cells to radiation involves autophagy and formation of acidic vesicles. *Cancer Res.* **61**, 439–444 (2001).
- Mizushima, N., Yamamoto, A., Matsui, M., Yoshimori, T. & Ohsumi, Y. *In vivo* analysis of autophagy in response to nutrient starvation using transgenic mice expressing a fluorescent autophagosome marker. *Mol. Biol. Cell* **15**, 1101–1111 (2004).
- Copp, A. J. Neurulation in the cranial region — normal and abnormal. *J. Anat.* **207**, 623–635 (2005).
- Liang, C. *et al.* Autophagic and tumour suppressor activity of a novel Beclin1-binding protein UVRAG. *Nature Cell Biol.* **8**, 688–698 (2006).
- Pattingre, S. *et al.* Bcl-2 antiapoptotic proteins inhibit Beclin 1-dependent autophagy. *Cell* **122**, 927–939 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Torres and P. Bonaldo for their involvement in the large-scale gene-trap screening. We also thank A. Conrad for the mouse work, S. Hille for DNA sequencing, M. Acuña Villa and M. W. Bennett for editorial and secretarial work, P. Mattioli for help with image processing, and G. Lisi, and G. Bellavia G. Marchetti for research assistance. We are grateful to N. Mizushima for providing us with the GFP-LC3 mice. This work was supported in part by grants from the Telethon Foundation and Compagnia di San Paolo (F.C.), Ricerca Corrente and Ricerca Finalizzata from Ministero della Salute and AIRC (M.P.) and the Max Planck Society (A.S., K.C. and P.G.). We thank the Ministry of University and Research of Italy for financial support.

Author Contributions G.M.F., A.R., S.D.B., C.F. and M.C. performed the protein interaction and autophagy assays. A.S. and K.C. generated the *Ambra1* gene-trap line and identified the tagged gene. A.S., L.G., K.C., R.N. and S.D.B. carried out the analysis of the phenotype. R.N. and P.S. performed electron microscopy. A.U. generated the *Ambra1*^{gst} fusion construct. P.G.'s laboratory devised and performed the large-scale gene-trap screening. F.C., M.P. and G.M.F. wrote the paper. All authors discussed the results and commented on the manuscript.

Author Information The sequence for human *AMBRA1* cDNA (Supplementary Fig. 1c) has been submitted to GenBank under accession number DQ870924. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to F.C. (francesco.cecconi@uniroma2.it) or K.C. (kchowdh@gwdg.de).

METHODS

Chromosomal localization. For chromosomal localization of the gene-trap insertion site, chromosomal slides were prepared from mouse spleen lymphocytes and hybridized with a biotinylated 2.7-kilobase (kb) cDNA insert as described in ref. 31.

Characterization of the *Ambra1*^{gt} mutation. Several subsequent rounds of 5' PCR with reverse transcription and cDNA library screenings were performed to obtain the 4,635 base pairs (bp) of compiled cDNA sequence. In detail, after anchored (5' rapid amplification of cDNA ends) PCR (ref. 32), the cDNA template obtained from embryonic stem cell mRNA was amplified by two subsequent rounds of nested PCR. Primers were derived from the *lacZ* sequence. The amplified products were then blotted and hybridized with a probe specific for the vector splice acceptor site. Once isolated, the corresponding fragment was used as a probe to screen an embryonic mouse cDNA library. After identification of the corresponding cDNA in the databases, the insertion site was localized in the largest 50.36-kb intron located between the 100- and 113-bp coding exons, as shown in Supplementary Fig. 1b. The vector insertion disrupted the protein at amino acid position 841 (sequence *Mus musculus* BAE33303.1), thus deleting 459 amino acids from the C-terminal end. For *Ambra1* mRNA analysis, poly(A)⁺ RNA was prepared from embryos and 2 µg were analysed by standard northern blot analysis.

The construct *Ambra1*^{gt}, encoding the Ambra1–LacZ fusion protein present in the gene-trap line, was generated as follows. Briefly, a construct containing the sequences encoding the N-terminal part of the murine Ambra1 protein as present in the gene-trap line (PCR form cDNA clone IMAGE:5400613) was inserted in-frame within a LacZ–pA (LacZ–polyadenylation site) vector. Next, the Ambra1–LacZ–pA fragment (6.1 kb) was isolated by *SmaI*–*NotI* digestion and cloned into the *SmaI*–*NotI* cut pCMV Sport 6 vector (German Genome Resource Centre, RZPD) in the sense orientation. To check whether the *Ambra1*^{gt} construct was able to generate the expected Ambra1–LacZ fusion protein in an eukaryotic cell line, we transfected it into the HEK293 cell line using the calcium phosphate method. Cells were fixed and stained for β-gal activity 24 h after transfection. Blue staining was already visible after 2 h. Positive and negative control plasmids were also included in the assay, producing the expected results. The full-length murine *Ambra1* clone (*Ambra1*^{wt}), used as a positive control, contained the entire coding sequence inserted into the pCMV Sport 6 vector in sense orientation.

Determination of genotype of *Ambra1* mutant embryos. To genotype embryos, a Southern blot analysis was performed after digestion of genomic DNA with the *Bam*HI restriction enzyme. The blots were incubated with radioactive probes specific for *lacZ* and for the housekeeping *Fkh5* gene. Densitometric comparison between the two signals was performed to discriminate between heterozygous and homozygous embryos. Early-stage embryos were genotyped by PCR with reverse transcription. Two pairs of primers were used: first, 5'-AACGCATTTATACCCAGTCCA-3' (primer A) and 5'-ACCATAACGTATCGGCCATC-3' (primer B), mapping upstream and downstream of the gene-trap insertion site, respectively; and, second, primer A together with 5'-CCCAGTCACGACGTTGTAAAA-3' (primer C), the latter one mapping on the *lacZ* reporter sequence.

Histological and immunocytochemical analyses. Whole-mount β-gal staining of embryos at different embryonic stages was performed as described³³. In brief, once removed from the uterus, embryos were washed in PBS and fixed in 4% paraformaldehyde (PFA) for 30 min at 4 °C. After washing, they were incubated at 30 °C in a staining solution containing 5-bromo-4-chloro-3-indolyl-β-D-galactoside for up to 15 h. The stained embryos were then cut with a vibratome into 60-µm-thick sections and were examined under a Zeiss Axioplan microscope.

For GFP–LC3 detection of *Ambra1*^{gt/gt};GFP–LC3 embryos, 10-µm-thick cryostat sections were prepared and analysed under an Olympus IX70 microscope using softWoRx and DeltaVision Imaging Workstation (Applied Precision).

Paraffin sectioning specimens were impregnated with paraffin wax (Paraplast Plus), embedded, transversally sectioned at 10 µm, and counterstained with cresyl-violet.

For immunohistochemical analysis, the *Ambra1*^{gt/gt} and wild-type embryos at E9–E9.5 were fixed in 4% PFA, washed and incubated in 25% sucrose, before cryo-embedding. Sagittal and coronal 10-µm-thick sections were blocked in 5% horse serum and then incubated overnight with the primary antibodies. Fluorescent secondary antibodies (Molecular Probes) were used for signal detection. Cellular nuclei were counterstained with 4,6-diamidino-2-phenylindole or propidium iodide (Molecular Probes). Sections were then examined under an inverted fluorescence microscope (Nikon and Leica). For ubiquitin detection, a biotinylated goat anti-mouse IgG was used as a secondary antibody followed by incubation with horseradish peroxidase-conjugated streptavidin (Biogenex). The immunoreaction product was revealed using 3-amino-9-ethyl-carbazole

as a chromogenic substrate and 0.01% H₂O₂ (Biogenex). In cell death experiments, apoptotic cells in embryonic sections were detected by a DNA fragmentation (TUNEL) assay using a fluorescein-based detection kit (Promega). Analysis of cell proliferation was performed after intraperitoneal injection of BrdU (Sigma–Aldrich) using an anti-BrdU antibody (Amersham). For confocal analysis of Becn1–Ambra1 co-localization and GFP–LC3-transduction experiments, cells were cultured on coverslips and fixed with 4% PFA in PBS followed by permeabilization with 0.1% Triton X-100 in PBS. Primary antibodies were incubated for 1 h at room temperature and visualized by means of Cy3- or Cy2-conjugated secondary antibodies (Jackson ImmunoResearch). Coverslips were mounted in Vectashield (Vector Laboratories) and examined under a confocal microscope (Leica TCS SP2).

Antibodies. The primary antibodies used in this study were: rabbit polyclonal anti-Myc Tag antibody (Upstate Biotechnology), goat polyclonal anti-Becn1 (Santa Cruz), mouse monoclonal anti-Becn1 (Becton–Dickinson), rabbit polyclonal anti-Vps34 (Invitrogen), rabbit polyclonal anti-phospho-p70S6K (Cell Signaling), mouse monoclonal anti-tubulin (Sigma–Aldrich), rabbit polyclonal anti-LC3 (MBL), mouse monoclonal anti β-gal (Promega), rabbit polyclonal anti-Shh (Santa Cruz) and mouse monoclonal anti-ubiquitin (1B3, MBL). The antibody against Ambra1 was raised by immunizing rabbits with a peptide located at the amino-terminal of Ambra1 protein (WEGKRVLPDSRSC) (Sigma Genosys) and was affinity-purified using the immunizing peptide.

Yeast two-hybrid screening. pGBKT7–AMBRA1 was generated by cloning the N-terminal region (nucleotides 1–2,001) of AMBRA1 into the *Eco*RI and *Bgl*II sites of pGBKT7 (Clontech). A human brain cDNA library cloned in pACT2 (Clontech) was screened by co-transformation with the pGBKT7–AMBRA1 into the AH109 yeast strain. Positive clones were selected based on their growth on Trp, Leu, Ade and His dropout media (Clontech) containing 5 mM 3-amino-1,2,4-triazole (Sigma–Aldrich). Recovery of the plasmids and the β-gal assay was performed following the manufacturer's instructions (Matchmaker Two-Hybrid System Protocol, Clontech).

Cell culture. The human fibrosarcoma 2FTGH cell line, provided by S. Pellegrini, and the human embryonic kidney HEK293 cells, provided by M. Pando, were cultured in DMEM (Sigma–Aldrich) supplemented with 10% FCS (Sigma–Aldrich), 2 mM L-glutamine, and 1% penicillin/streptomycin solution at 37 °C under 5% CO₂. For autophagy induction, cells were treated with 2 µM rapamycin (Sigma–Aldrich) or cultured in Earle's balanced salt solution. MEF primary cells were prepared from E13.5 embryos, cultured in DMEM supplemented with 20% FCS, and utilized for experiments between the second and the seventh passages.

RNA interference. siRNA oligoribonucleotides corresponding to the human AMBRA1 and BECN1 cDNA sequences were purchased from Dharmacon (AMBRA1) and Invitrogen (BECN1, BECN1 Stealth Select RNAi; HSS112742). AMBRA1 siRNA 1: 5'-AGAACTGCAAGATCTACAA-3'; AMBRA1 siRNA 2: 5'-GGCCTATGGTACTAACAAA-3'.

A total of 2 × 10⁵ cells per well were transfected with 100 pmol siRNA in six-well plates by Lipofectamine 2,000 (Invitrogen), as indicated by the supplier. Transfection was repeated on two consecutive days to increase transfection efficiency. Twenty-four hours after transfection, cells were digested with trypsin, plated at 5 × 10⁴ cells per well in six-well plates and treated with rapamycin (Sigma–Aldrich). The RNA decrease was checked by real-time PCR and western blot 48 h after transfection.

Cell proliferation analysis. To measure the percentage of cells entering in S phase, BrdU was added to a final concentration of 20 µM and incubated at 37 °C for 20 min. Cells were detached using trypsin, washed with PBS and fixed using cold 70% ethanol. DNA was denatured by adding 4 N HCl for 1 h at room temperature, followed by neutralization with 0.5 M Na₂B₄O₇ (pH 8.5). Cells were then incubated with anti-BrdU Alexa Fluor 488 (0.2 µg ml⁻¹, Molecular Probes) in PBS plus 0.5% Tween 20 for 1 h at 37 °C. Finally, cells were washed with PBS, resuspended in PBS plus 5 µg ml⁻¹ propidium iodide and analysed using a FACScan flow cytometer (Becton Dickinson).

cDNA cloning and retroviral vectors. For retroviral expression, all constructs were cloned in a modified version of pLPCX vector (Clontech) in which the *Sal*I restriction site within the puromycin-resistance gene was mutated by PCR to render the *Sal*I site unique in the multiple cloning site. An additional cytomegalovirus promoter was cloned in front of the 5' long terminal repeat region to increase viral production³⁴.

Human AMBRA1 cDNA (Supplementary Fig. 1c) was obtained by PCR amplification from a human brain cDNA library (Clontech) and cloned in *Eco*RI and *Not*I sites of the pLPCX vector. The various pLPCX–AMBRA1 mutants (NT, CT, F1, F2, F3, F4 and F5) were generated by PCR amplification using specific primers. Myc-tagged AMBRA1 fusion proteins were obtained by inserting five copies of the Myc epitope in the *Hind*III–*Eco*RI restriction sites of pLPCX. Human BECN1 cDNA was obtained by PCR amplification from a HeLa

cDNA library (Clontech) and cloned into the *EcoRI* and *Sall* sites of pCLPCX. Human LC3 cDNA was obtained by PCR amplification from a HeLa cDNA library (Clontech) and inserted into the *EcoRI* and *Sall* sites of pEGFP/C2 (Clontech). GFP-LC3 fusion cDNA was then excised from the pEGFP plasmid by cutting with *NheI* and *Sall* restriction enzymes and inserted into pLPCX digested with *EcoRI* and *Sall*, with *NheI* and *EcoRI* DNA ends blunted to allow the ligation. pLPCX-lacZ-myc was obtained by digesting the pcDNA 3.1 myc-His/lacZ plasmid (Invitrogen) with *HindIII*-*PmeI* and inserted into pLPCX digested with *HindIII*-*NoI* (blunted). Human p40(phox)PX cDNA (encoding amino acids 13–140 of human p40phox protein) was obtained by PCR amplification from a HeLa cDNA library (Clontech), inserted into the *EcoRI* and *Sall* sites of pEGFP/C2 (Clontech) and transfected into the HEK293 cell line using the calcium phosphate method. The sequences of all PCR-amplified cDNAs were verified by DNA sequencing analysis.

Retrovirus generation and infection. Fifteen micrograms of the retroviral vectors were co-transfected with 5 µg of an expression plasmid for the vesicular stomatitis virus G protein into the cell line 293 gp/bsr (ref. 35) using the calcium phosphate method. Forty-eight hours later, the supernatant containing the retroviral particles was recovered and supplemented with polybrene (4 µg ml⁻¹). 2FTGH or MEF cells were infected by incubation with retroviral-containing supernatant for 6–8 h.

Real-time PCR. RNA was prepared with Trizol reagent (Invitrogen). cDNA synthesis was generated using the reverse transcription kit (Promega) according to manufacturer's recommendations. Real-time PCR reactions were performed with the LightCycler (Roche). The LightCycler FastStart DNA Master SYBR Green I (Roche) was used to produce fluorescent-labelled PCR products during repetitive cycling of the amplification reaction. Primer sets for all amplicons were designed using the Primer-Express 1.0 software system. *AMBRA1* forward: 5'-AACCCTCCACTGCGAGTTGA-3'; *AMBRA1* reverse: 5'-TCTACCTGTTCCGTGGTCTCTCC-3'; *L34* forward: 5'-GTCCCGAACCCTGGTAATAGA-3'; *L34* reverse: 5'-GGCCCTGCTGACATGTTTCTT-3'.

The result of the fluorescent PCR was expressed as the threshold cycle (C_T). The ΔC_T is the difference between the C_T for a specific mRNA and the C_T for a reference mRNA, *L34*. To determine relative mRNA levels, 2 was raised to the power of $\Delta\Delta C_T$ (the difference between the ΔC_T from treated cells and the C_T from untreated cells). *L34* mRNA level was used as an internal control because this gene was shown to be stable with cell induction. β -actin and glyceraldehyde-3-phosphate dehydrogenase levels were used as additional controls to confirm significant decreases.

Immunoprecipitation and western blot assays. In immunoprecipitation experiments, cells or tissues were lysed in HEMG buffer (25 mM HEPES

(pH 8.0), 100 mM NaCl, 0.5% Nonidet P-40, 0.1 mM EDTA, 10% glycerol) plus protease and phosphatase inhibitors (protease inhibitor cocktail, 1 mM sodium fluoride, 1 mM sodium orthovanadate, 1 mM sodium molybdate; Sigma-Aldrich). Lysates (1–2 mg) were then incubated with rotation at 4 °C for 30 min. After a centrifugation at 4 °C for 10 min at 13,000g to remove insoluble debris, equal amounts of protein were incubated with 30 µl monoclonal anti-cMyc antibody conjugated with protein A agarose beads (BD Biosciences) with rotation at 4 °C for 4 h, or with 2 µg of BECN1 or AMBRA1 antibodies overnight at 4 °C, followed by 60 min incubation with 20 µl of protein A/G sepharose beads (Amersham Bioscience). The beads were finally collected by centrifugation and washed four times with the HEMG buffer. Proteins bound to the beads were eluted with 50 µl of SDS-polyacrylamide gel electrophoresis sample buffer and heated to 70 °C for 10 min.

Proteins were separated on NuPAGE Bis-Tris gel (Invitrogen) and electro-blotted onto nitrocellulose membranes (Protran, Schleicher and Schuell). Blots were incubated with primary antibodies in 5% non-fat dry milk in PBS plus 0.1% Tween 20 overnight at 4 °C. Detection was achieved using horseradish peroxidase-conjugate secondary antibody (Jackson Laboratory) and visualized with ECL plus (Amersham Bioscience).

Embryos from stage E11–E11.5 to E14–E14.5 were subjected to mechanical lysis in 50 mM Tris HCl (pH 7.5), 320 mM sucrose, 50 mM NaCl, 1% Triton X-100 and protease inhibitors. Solubilized proteins were quantified by a Biorad protein assay, and denatured by adding a concentrated boiling Laemmli buffer. A total of 30–50 µg of protein was analysed by SDS-polyacrylamide gel electrophoresis and immunoblotting.

Statistical analysis. Microsoft Excel was used for statistical analysis. Statistical significance was determined using the Student's *t*-test. A *P* value of less than or equal to 0.05 was considered significant.

31. Heng, H. H. & Tsui, L. C. Modes of DAPI banding and simultaneous *in situ* hybridization. *Chromosoma* **102**, 325–332 (1993).
32. Chowdhury, K., Bonaldo, P., Torres, M., Stoykova, A. & Gruss, P. Evidence for the stochastic integration of gene trap vectors into the mouse germline. *Nucleic Acids Res.* **25**, 1531–1536 (1997).
33. Stoykova, A., Chowdhury, K., Bonaldo, P., Torres, M. & Gruss, P. Gene trap expression and mutational analysis for genes involved in the development of the mammalian nervous system. *Dev. Dyn.* **212**, 198–213 (1998).
34. Ranga, U. *et al.* Cell and viral regulatory elements enhance the expression and function of a human immunodeficiency virus inhibitory gene. *J. Virol.* **71**, 7020–7029 (1997).
35. Somia, N. V. *et al.* LFG: an anti-apoptotic gene that provides protection from Fas-mediated cell death. *Proc. Natl. Acad. Sci. USA* **96**, 12667–12672 (1999).

LETTERS

miRNAs control gene expression in the single-cell alga *Chlamydomonas reinhardtii*

Attila Molnár¹, Frank Schwach^{1,2}, David J. Studholme¹, Eva C. Thuenemann¹ & David C. Baulcombe¹

MicroRNAs (miRNAs) in eukaryotes guide post-transcriptional regulation by means of targeted RNA degradation and translational arrest¹. They are released by a Dicer nuclease as a 21–24-nucleotide RNA duplex from a precursor in which an imperfectly matched inverted repeat forms a partly double-stranded region. One of the two strands is then recruited by an Argonaute nuclease that is the effector protein of the silencing mechanism. Short interfering RNAs (siRNAs), which are similar to miRNAs, are also produced by Dicer but the precursors are perfectly double-stranded RNA. These siRNAs guide post-transcriptional regulation, as with miRNAs, and epigenetic genome modification. Diverse eukaryotes including fungi, plants, protozoans and metazoans produce siRNAs^{2–5} but, until now, miRNAs have not been described in unicellular organisms and it has been suggested that they evolved together with multicellularity in separate plant and animal lineages⁶. Here we show that the unicellular alga *Chlamydomonas reinhardtii* contains miRNAs, putative evolutionary precursors of miRNAs and species of siRNAs resembling those in higher plants. The common features of miRNAs and siRNAs in an alga and in higher plants indicate that complex RNA-silencing systems evolved before multicellularity and were a feature of primitive eukaryotic cells.

The *Chlamydomonas* genome encodes both Dicer and Argonaute nuclease (AGO) proteins, indicating that it is fully competent to perform RNA silencing⁷. To characterize RNA silencing in this alga we generated 46,565 sequences of small silencing RNAs (sRNAs) 2–32 nucleotides (nt) in size. After the removal of sequences shorter than 16 nt or matching transfer RNAs or ribosomal RNAs, 26,427 reads (57% of total) had at least one exact match on the *Chlamydomonas* genome (Assembly v.3.0; <http://genome.jgi-psf.org/Chlre3/Chlre3.home.html>). Most sequences fell into the 20–22-nt size class; the abundant 24-nt size class of *Arabidopsis* siRNAs⁸ was absent

(Fig. 1a). Many sequences matched the genome multiple times, resulting in 257,156 genome matches. Similarly to sRNAs in other species, the *Chlamydomonas* sRNAs display a bias towards A or U nucleotides at the 5' end (Fig. 1b)^{9,10}.

We identified candidate genetic loci with the potential to produce siRNAs and miRNAs by searching for regions of the genome with multiple sRNA matches (see Methods). As a result of redundancy in the draft genome assembly, these loci were grouped by nucleotide sequence similarity and further analysis was restricted to one representative locus in each group. The representative loci accounted for 96% of genome-matching sequence reads.

Of 1,479 representative loci there are 207 (14%) with sequence similarity to transposable elements including long-terminal-repeat retrotransposons (Supplementary Table 1) (ftp://ftp.jgi-psf.org/pub/JGI_data/Chlamy/v3.0/CHLREP.fn.gz). We can exclude the possibility that these transposon-specific RNAs were random degradation products because they are predominantly of the 20–22-nt sRNA class, and transposon sequences are over-represented relative to the sequenced genome (Supplementary Fig. 1). These loci are therefore likely to encode siRNAs involved in genome defence.

There are also four *Chlamydomonas* loci at which the alignment of sRNA sequence reads was phased (Supplementary Table 2 and Supplementary Fig. 2). RNA folding algorithms did not predict inverted repeat regions in the transcripts of these loci, indicating that they might be similar to *trans*-acting siRNA¹¹ and natural antisense siRNA loci in *Arabidopsis*¹². The phasing analyses revealed that a high proportion of the siRNAs matching these loci were in one or two registers (Fig. 2a, Supplementary Table 3 and Supplementary Fig. 3). Northern hybridization for the two loci predicted to produce abundant siRNAs confirmed that these siRNAs were present in light-grown and dark-grown cells. The siRNA of locus 2689 (CrsRNA2) accumulated preferentially in gametes, whereas CrsRNA26 was induced by light (Fig. 2b).

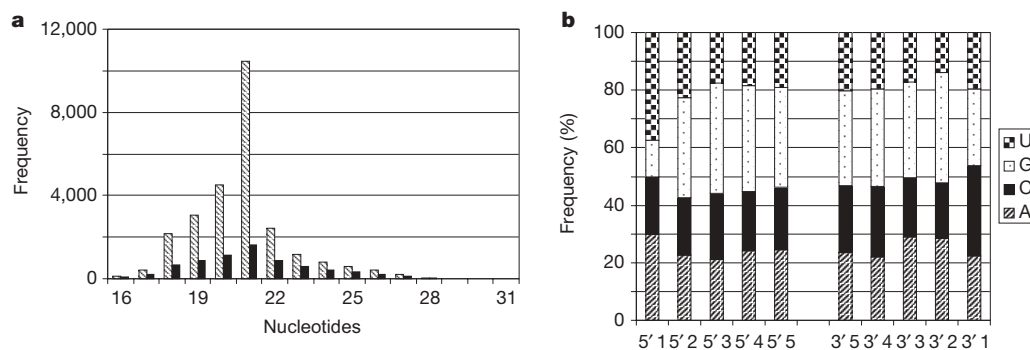


Figure 1 | *Chlamydomonas* sRNAs. **a**, Size distribution of *Chlamydomonas* sRNAs. Redundant and non-redundant sRNA reads are represented as hashed and solid bars, respectively. **b**, Nucleotide frequency at the 5' and 3'

end of non-redundant sRNAs. The nucleotide positions are numbered from the 5' and 3' ends of sequence reads as indicated.

¹Sainsbury Laboratory, John Innes Centre, Norwich NR4 7UH, UK. ²University of East Anglia, School of Computing Sciences, Norwich NR4 7TJ, UK.

The phased siRNAs must derive from double-stranded RNA (dsRNA) because they correspond to both RNA strands (Supplementary Table 2). In *Arabidopsis* the biogenesis of phased *trans*-acting siRNAs and natural antisense siRNAs involves the conversion of a single-stranded RNA precursor into a double-stranded form by an RNA-dependent RNA polymerase^{12,13} (RdRP). However, there is no recognizable homologue of the RdRP encoded in the *Chlamydomonas* genome⁷. Perhaps there is RNA templated activity of DNA-dependent RNA polymerases or bidirectional transcription of DNA at the phased RNA loci and annealing of the complementary RNAs. In addition, because of the incomplete nature of the *Chlamydomonas* genome, we cannot exclude the possibility that some of these loci generate foldback transcripts that are cleaved directly by Dicer to generate the phased siRNAs.

To search for potential *Chlamydomonas* miRNA loci we screened computationally for the ability of long transcripts from sRNA loci to fold into precursor miRNA-like molecules with imperfectly matched inverted repeats. The candidate miRNA loci were selected as those in which 80% or more of the sRNAs matched hairpin regions and in which the secondary structure of the predicted transcript was significantly ($P < 0.05$) more stable than that of the randomized sequence. A full list of the 68 candidate miRNA loci returned by this algorithm is given in Supplementary Table 4; the miRNA sequences and the precursor structures can be found in Supplementary Tables 5 and 6 and Supplementary Figs 4 and 5.

We subdivided the 68 candidate miRNA loci into two categories. The first 'short hairpin' category (21 out of 68) resembled miRNA loci of higher plants and animals¹ in that the hairpin regions were shorter than 150 nt and they specified a single miRNA (Fig. 3a, top four panels). In 12 instances, the opposite strand of the sRNA duplex termed 'miRNA*' was represented, although less frequently than the miRNA. A genome browser view of six loci is shown in Supplementary Fig. 6. The putative miRNA sequence at these loci was typically located 18–24 nt away from the tip of the predicted hairpin RNA, suggesting the importance of terminal structures in the recognition of miRNA precursors by the *Chlamydomonas* Dicer complex. This feature is not found in higher-plant or animal miRNA precursors. Animal miRNA-processing enzymes use structural cues at the base of the dsRNA stem¹⁴.

The predicted transcripts of remaining miRNA-like loci in *Chlamydomonas* can form long (150–729 nt), almost perfect hairpins, with the potential to produce multiple sRNAs (Fig. 3a, bottom two

panels). The small RNAs on three of these long hairpins showed phasing, as with the loci referred to above (Supplementary Table 4). The sRNAs are absent from the predicted loop region and, as with the short hairpin loci, the first sRNA was located 18–24 nt from the terminal loop structure. It is likely that these loci represent young miRNA genes giving rise to multiple sRNAs.

These miRNA and miRNA-like RNAs are expressed and, in eight examples tested, we could detect them by northern blotting (Fig. 3a, b). In most instances one strand—the presumed miRNA strand—was more abundant than the miRNA* strand (Fig. 3a, left panel). None of these putative miRNAs exhibited light-dependent expression, but long hairpin-derived sRNAs 61 and 248 and short hairpin-derived sRNAs 601 and 2726 were upregulated in gametes (Fig. 3b).

The miRNAs in higher plants direct site-specific cleavage of their target RNA, and in many instances the 3' cleavage fragment of the target mRNA is stable and can be detected by a modified 5' rapid amplification of cDNA ends (RACE) assay¹⁵. To find out whether a specific cleavage of miRNA target RNA occurs in *Chlamydomonas*, we first applied the *Arabidopsis* rules for target prediction¹⁶ to the putative miRNA from the long and short hairpin precursors (Supplementary Tables 5 and 6). Using RNA from vegetative cells and 5' RACE to detect the predicted 3' cleavage products, we obtained a single polymerase chain reaction (PCR) product of the size expected for 4 of 18 putative miRNA targets tested. These examples included putative miRNAs of both short (cr.01847, cr.01848 and cr.02386) and long (cr.02573) loci and, in each instance, the most frequent cleavage site was at the centre of the miRNA target site, which is consistent with the prediction of AGO cleavage. It is therefore likely that the AGO-mediated target-RNA cleavage takes place in *Chlamydomonas*, as in *Arabidopsis* and other higher plants (Fig. 4a).

Among the verified miRNA targets were the mRNAs of two flagellum-associated proteins (protein IDs 159889 and 183179)¹⁷ and a potential NAD⁺-dependent sugar epimerase/dehydratase (protein ID 193550). Failure to verify miRNA targets with the other 14 putative miRNAs experimentally could be related to the as yet incomplete annotation of the *Chlamydomonas* genome and the possibility that the potential sRNA targets are not expressed in the vegetative cells analysed by us. It is also possible that the miRNA cleavage products are not stable enough for identification by 5' RACE.

To explore further the similarity of silencing in *Arabidopsis* and *Chlamydomonas* we investigated 3' modifications of the *Chlamydomonas* miRNAs and siRNAs. In *Arabidopsis* the sRNAs are

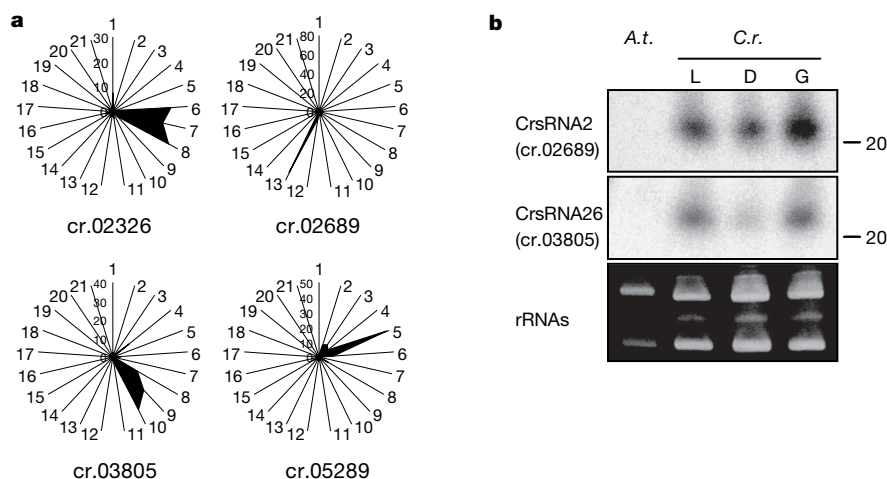


Figure 2 | Phasing siRNA loci in *Chlamydomonas*. **a**, Phasing registers. Each spoke of the radial graph indicates the percentage of 21-nt sRNA matches belonging to one of the 21 possible phasing registers. Locus lengths were as follows: cr.02326, 787 base pairs (bp); cr.02689, 289 bp; cr.03805, 363 bp; cr.05289, 178 bp. **b**, Northern analyses of phased siRNAs. CrsRNA2 of locus cr.02689 is the most abundant phasing group on the reverse strand in

register 13; CrsRNA26 of locus cr.03805 is the second most abundant phasing siRNA on the reverse strand in register 9. RNA was prepared from cell cultures grown in the light (L), in the dark (D) and from gametes (G). rRNAs stained with ethidium bromide are shown as a loading control. *A. t.*, *Arabidopsis thaliana*; *C. r.*, *Chlamydomonas reinhardtii*.

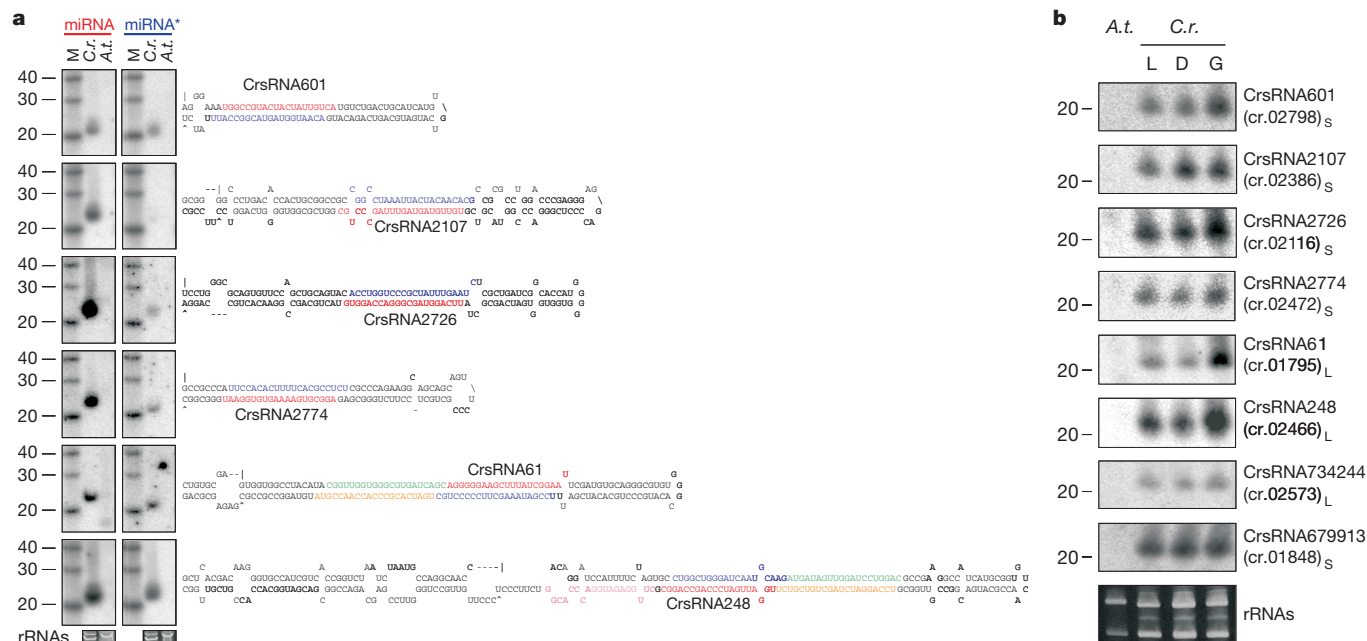


Figure 3 | *Chlamydomonas* miRNAs. **a**, Detection of miRNA and miRNA* by small-RNA northern hybridization, and representative secondary structures of miRNA precursors. miRNAs and miRNA*s are highlighted in red or green and in blue or orange, respectively. CrsRNA248* was not sequenced from our sRNA cDNA library; however we were able to detect it by small-RNA northern hybridization. **b**, Expression profile analyses of eight

different *Chlamydomonas* miRNAs by northern hybridization. sRNAs derived from short (S) and long (L) predicted miRNA precursors are indicated. Cluster numbers are shown in brackets. Other abbreviations are as in Fig. 2. Genome browser views of miRNA loci are shown in Supplementary Fig. 6.

methylated by HEN1 (ref. 18) at the 2' hydroxy position of the 3' nucleotide¹⁹. This modification protects the 3' ends from uridylation and degradation²⁰. *Chlamydomonas* siRNAs and miRNAs are likely to

have the same 3' modification because, like *Arabidopsis* miRNA and unlike a synthetic siRNA, they were resistant to degradation by β -elimination (Fig. 4b). A predicted *Chlamydomonas* protein

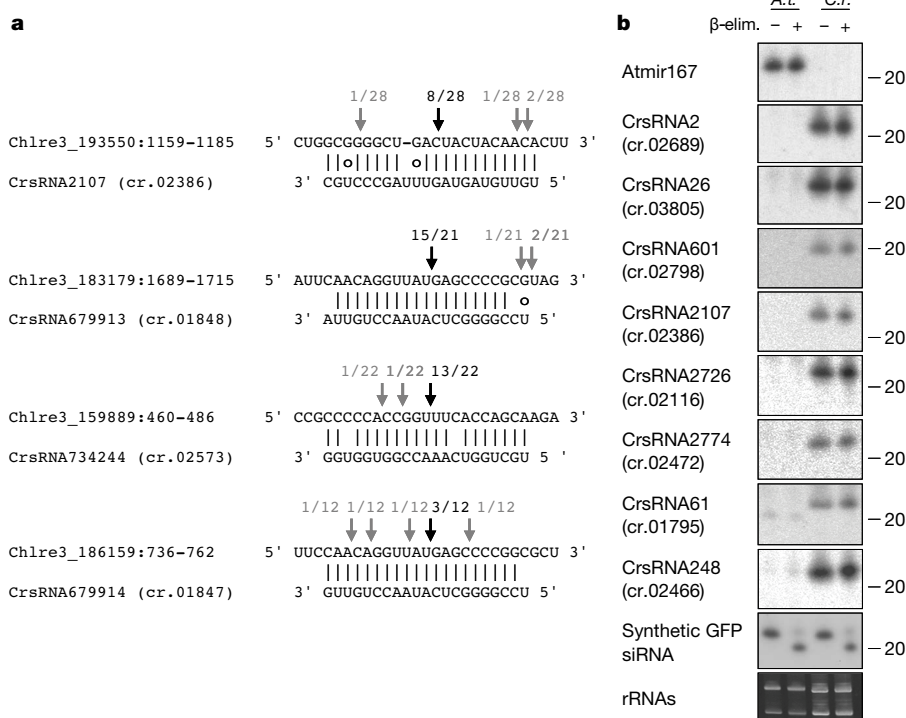


Figure 4 | miRNA action and protection. **a**, Verification of predicted miRNA targets by 5' RACE PCR. Arrows and numbers indicate the positions and the frequencies of 5' termini of truncated target mRNAs (top) aligned with miRNAs (bottom). Circles indicate G:U wobble pairing. **b**, Terminal structure of *Chlamydomonas* siRNAs and miRNAs. Total RNAs mixed with synthetic GFP siRNAs were subjected to RNA filter hybridization with

probes against various siRNAs and miRNAs. Cluster numbers are shown in brackets. Increased mobility of small RNAs in treated samples (β -elim. +) indicates unprotected, terminal hydroxyl groups. rRNAs stained with ethidium bromide are shown as a loading control. A.t., *Arabidopsis thaliana*; C.r., *Chlamydomonas reinhardtii*.

(Chlre3_170680) is similar to *Arabidopsis* HEN1 in that it has both a Rossmann-fold methyltransferase (RFM) domain and a pair of putative dsRNA-binding motifs (DSRMs) (Supplementary Fig. 7).

These combined findings about miRNA precursors, miRNA targeting and 3'-end modification show that miRNAs in *Chlamydomonas* are similar in many respects to those in higher plants. However, we did not detect any miRNAs with direct sequence similarity between algae and either higher plants or animals (data not shown). In contrast, in mosses and other primitive land plants, there are some miRNAs that are almost identical to those in angiosperms^{21,22}. We propose that the absence of conservation of sequences with higher plants is because miRNAs in algae are less mature in an evolutionary sense than those in higher plants. Consistent with this idea is the observation that the miRNA loci in *Arabidopsis* are predominantly like the short hairpin loci described here, with the characteristics of an endpoint in miRNA evolution⁶. In contrast, in *Chlamydomonas*, the long hairpin loci with the characteristics of an early stage in miRNA evolution were the more prevalent form (47 out of 68 loci).

RNA silencing is a part of defence systems against viral and transposon invaders. It is also involved in regulatory mechanisms, with the potential to act at either the transcriptional or the post-transcriptional level. From the similarities of RNA silencing in *Chlamydomonas* and *Arabidopsis* it is likely that diverse RNA silencing mechanisms were present in their common ancestor—presumably a primitive unicellular alga. Particularly surprising is the finding that both organisms and, presumably, their common ancestor, have a specialized type of silencing RNA—miRNAs—that had been previously associated with developmental regulation and multicellularity. This alga can now be used to analyse complex systems involving positive and negative feedback mechanisms that are a frequent feature of regulation by siRNAs and miRNAs²³.

METHODS SUMMARY

Chlamydomonas reinhardtii was grown in liquid culture and total RNA was extracted. A cDNA library was produced by PCR amplification of small-RNA cDNA, and end-labelled oligonucleotide probes were used for northern blotting. β -elimination was performed to detect 3' modification of RNA. 5' RACE on poly(A)-selected RNA was used to detect miRNA cleavage products. Computational analysis of small RNA data was based on reads with sequence identity to the *Chlamydomonas* genome.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 21 February; accepted 2 May 2007.

Published online 30 May 2007.

1. Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281–297 (2004).
2. Reinhart, B. J. & Bartel, D. P. Small RNAs correspond to centromere heterochromatic repeats. *Science* **297**, 1831 (2002).
3. Llave, C., Kasschau, K. D., Rector, M. A. & Carrington, J. C. Endogenous and silencing-associated small RNAs in plants. *Plant Cell* **14**, 1605–1619 (2002).
4. Dijkeng, A., Shi, H. F., Tschudi, C. & Ullu, E. RNA interference in *Trypanosoma brucei*: Cloning of small interfering RNAs provides evidence for retroposon-derived 24–26-nucleotide RNAs. *RNA* **7**, 1522–1530 (2001).

5. Ambros, V., Lee, R. C., Lavanway, A., Williams, P. T. & Jewell, D. MicroRNAs and other tiny endogenous RNAs in *C. elegans*. *Curr. Biol.* **13**, 807–818 (2003).
6. Allen, E. et al. Evolution of microRNA genes by inverted duplication of target gene sequences in *Arabidopsis thaliana*. *Nature Genet.* **36**, 1282–1290 (2004).
7. Cerutti, H. & Casas-Mollano, J. A. On the origin and functions of RNA-mediated silencing: from protists to man. *Curr. Genet.* **50**, 81–99 (2006).
8. Lu, C. et al. MicroRNAs and other small RNAs enriched in the *Arabidopsis* RNA-dependent RNA polymerase-2 mutant. *Genome Res.* **16**, 1276–1288 (2006).
9. Schwarz, D. S. et al. Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115**, 199–208 (2003).
10. Khvorova, A., Reynolds, A. & Jayasena, S. D. Functional siRNAs and miRNAs exhibit strand bias. *Cell* **115**, 209–216 (2003).
11. Vazquez, F. et al. Endogenous trans-acting siRNAs regulate the accumulation of *Arabidopsis* mRNAs. *Mol. Cell* **16**, 69–79 (2004).
12. Borsani, O., Zhu, J., Verslues, P. E., Sunkar, R. & Zhu, J.-K. Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in *Arabidopsis*. *Cell* **123**, 1279–1291 (2005).
13. Peragine, A., Yoshikawa, M., Wu, G., Albrecht, H. L. & Poethig, R. S. SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of trans-acting siRNAs in *Arabidopsis*. *Genes Dev.* **18**, 2368–2379 (2004).
14. Han, J. et al. Molecular basis for the recognition of primary micro-RNAs by the Drosha-DGCR8 complex. *Cell* **125**, 887–901 (2006).
15. Llave, C., Xie, Z., Kasschau, K. D. & Carrington, J. C. Cleavage of *Scarecrow-like* mRNA targets directed by a class of *Arabidopsis* miRNA. *Science* **297**, 2053–2056 (2002).
16. Schwab, R. et al. Specific effects of microRNAs on the plant transcriptome. *Dev. Cell* **8**, 517–527 (2005).
17. Pazour, G. J., Agrin, N., Leszyk, J. & Witman, G. B. Proteomic analysis of a eukaryotic cilium. *J. Cell Biol.* **170**, 103–113 (2005).
18. Yu, B. et al. Methylation as a crucial step in plant microRNA biogenesis. *Science* **307**, 932–935 (2005).
19. Yang, Z., Ebright, Y. W., Yu, B. & Chen, X. HEN1 recognizes 21–24 nt small RNA duplexes and deposits a methyl group onto the 2' OH of the 3' terminal nucleotide. *Nucleic Acids Res.* **34**, 667–675 (2006).
20. Li, J., Yang, Z., Yu, B., Liu, J. & Chen, X. Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in *Arabidopsis*. *Curr. Biol.* **15**, 1501–1507 (2005).
21. Floyd, S. K. & Bowman, J. L. Ancient microRNA target sequences in plants. *Nature* **428**, 485–486 (2004).
22. Arazi, T. et al. Cloning and characterization of micro-RNAs from moss. *Plant J.* **43**, 837–848 (2005).
23. Baulcombe, D. RNA silencing in plants. *Nature* **431**, 356–363 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Purton for helpful discussions and technical advice, and E. Havecker and R. A. Mosher for critical comments on the manuscript. This work was supported by the Gatsby Charitable Foundation. A.M. was a recipient of a long-term EMBO fellowship. F.S. was supported in part by a BBSRC grant and E.C.T. was supported by a Marie Curie Early Stage Training Fellowship. The *Chlamydomonas* genome sequence data were produced by the US Department of Energy Joint Genome Institute (<http://www.jgi.doe.gov/>) and are provided for use in this publication only.

Author Contributions A.M. and D.C.B. designed research. A.M. and E.C.T. performed the experiments. F.S. and D.J.S. developed the sRNA pipeline and performed the computational analyses. A.M., F.S., D.J.S. and D.C.B. analysed the data. A.M., F.S. and D.C.B. wrote the paper.

Author Information The small RNA sequences have been deposited in GEO with the accession number GSE7575. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.C.B. (david.baulcombe@tsl.ac.uk).

METHODS

Strain and culture conditions. The *Chlamydomonas reinhardtii* strain J3(mt[−]) carries the cell-wall-deficient mutation cw15 and was obtained from S. Purton. Vegetative cells were grown in TAP medium at 20 °C in complete darkness or under continuous illumination (85 $\mu\text{mol m}^{-2} \text{s}^{-1}$). Gametogenesis was induced by transfer of mid-exponential phase (cell density $(1-2) \times 10^6 \text{ cells ml}^{-1}$) vegetative cultures to nitrogen-free medium and incubation for 21 h under continuous light²⁴.

Construction of small RNA cDNA libraries. Total RNA was extracted from vegetative and gametogenesis induced cells with TriReagent in accordance with the manufacturer's instructions (Sigma). The small-RNA cDNA library was generated from 500 μg of total RNA prepared from vegetative cells as described previously²⁵. For deep sequencing, the cDNAs were amplified with primers 454-for (5'-GCC TCC CTC GCG CCA TCA GCA GCC ATG GGA ATT CCT CAC TAA-3') and 454-revA (5'-GCC TTG CCA GCC CGC TCA GAC AGT CCA TGG ATT G-3'), and the PCR products were separated in 10% native polyacrylamide gel. After ethidium bromide staining the approximately 100-bp band corresponding to amplified ligation products was excised and extracted with 0.3 M NaCl.

Northern blot analyses. Total RNA (30–50 μg) was separated on a 15% denaturing polyacrylamide Tris/borate/EDTA gel; the gel was then soaked in 10 mM sodium phosphate buffer pH 7.0 and subsequently in $20 \times \text{SSC}$ for 10 min. The RNAs were transferred by overnight capillary blotting in $20 \times \text{SSC}$ to Zeta-probe (Bio-Rad) or Hybond N⁺ (Amersham Biosciences) membranes. Membranes were crosslinked and/or baked for 2 h at 80 °C. Probes complementary to sRNAs were end-labelled with [γ -³²P]ATP with the use of T4 polynucleotide kinase (New England BioLabs). Hybridization was performed in 5 ml of 7% SDS in 0.25 M sodium phosphate buffer pH 7.2, and the hybridization temperature was set 15 °C below the T_m of oligonucleotides. Membranes were washed twice at hybridization temperature with $2 \times \text{SSC}$, 0.1% SDS for 10 min, and exposed to PhosphorImage plates. Decade RNA marker was labelled in accordance with the manufacturer's instruction (Ambion).

Beta elimination. Synthetic green fluorescent protein RNA (5'-CGU GCU GAA GUC AAG UUU GAG-3'; 2 pmol; Dharmacon) was mixed with total RNA extracts and β -elimination was performed as described previously²⁶.

Validation of miRNA and siRNA targets. Target validation was performed by 5' RACE with the use of the GeneRacer Kit (Invitrogen) on poly(A)-selected RNA from vegetative cells and gametes, omitting the decapping step; nested PCR was performed for each gene. Gene-specific primers were designed about 350 nt to the 3' side of predicted target sites. PCR reactions were separated on a 1.5% TBE-agarose gel, and distinct bands of the appropriate size for sRNA-mediated cleavage products were purified, cloned and sequenced.

Computational analyses. After trimming adaptor sequences, pyrosequenced reads²⁷ were inserted into a custom database. Sequences fully matching tRNAs or rRNAs were separated and not included in the downstream analyses presented here. The remaining sequences were aligned to the *Chlamydomonas* genome version 3.0, consisting of 1,556 assembled genome sequence scaffolds and 278,284 unassembled bonus scaffolds (available from <http://genome.jgi-psf.org/Chlre3/Chlre3.home.html>). Only full-length identical matches were used for further analysis.

Identification of sRNA loci and representative loci. Using the data set of exact full-length genomic matches, we scanned the genome for regions with at least four sRNA matches and with no more than 300 bp between any two matches. The choice of a 300-bp maximum gap was based on an analysis of distances between sRNA matches (Supplementary Fig. 8). Because of the unfinished nature of the *Chlamydomonas* genome it is not possible to distinguish between truly repetitive loci and loci that are present on more than one genomic sequence fragment because of redundancy in the assembly. To overcome this problem, we clustered the loci according to sequence similarity with the BLASTCLUST software (available at <http://www.ncbi.nlm.nih.gov/blast/>) with a stringent cut-off of 98% sequence identity covering at least 90% of the sequence length. Each group was then represented by the longest sequence present on a scaffold. Bonus scaffold sequences were used only in cases where no scaffold sequence was available. Remaining redundancy between representative loci was analysed manually for the sets of miRNA and phased-siRNA candidate loci by BLASTCLUST and the MUSCLE²⁸ multiple sequence alignment software.

Characterization of loci. For each representative locus, we searched for overlaps with publicly available genome annotation data (Joint Genome Institute, <http://genome.jgi-psf.org/Chlre3/Chlre3.home.html>). In addition, we scanned the loci for sRNA matches in a 21-nt phase, indicative of sequential Dicer cleavages. Only sRNAs 21 nt long were accepted and there had to be at least three different positions occupied by sRNAs in a phasing register. Sequences on opposite strands were combined into a single register if there was a 2-nt shift that would result in the typical 3' overhang of an siRNA duplex.

Loci in which at least 20% of sequence reads could be aligned to one of the 21 possible phasing registers were further tested for significance by a randomization test in which the alignment of sRNA matches was randomly shuffled 1,000 times and z -scores ($z = (x - \mu)/\delta$, where μ is the mean, δ is the standard deviation of the random sample and x is the observed value) were obtained for the percentage of sRNAs allocated to phasing registers and the number of occupied registers. Loci with z -scores less than -2 for the number of phasing registers or greater than $+2$ for the percentage of sRNAs in phasing groups were marked as 'phased'.

Identification of miRNA candidate loci. Up to 11 sequence windows on both strands, spanning the locus and including variable lengths of flanking regions (5–250 bp on either side), were examined for their potential to form fold-back transcripts by using the RNAfold program²⁹. The sRNA matches from each locus were mapped onto all folded sequence windows. Sequences overlapping each other by more than 80% were combined into overlap groups. Groups more than 29 nt in length were discarded. For all downstream analyses, one sRNA was selected to represent each overlap group. On the basis of the assumption that sRNAs shorter than 20 nt could be degradation products, our algorithm selected the most abundant sRNA that was at least 20 nt long or, if no such sequence was present, the most abundant sequence shorter than 20 nt. If a decision could not be made on the basis of abundance, the algorithm would select the longest sRNA. The selected representative sRNA of each overlap group was accepted as a potential miRNA or miRNA* if it passed all of the following criteria derived from known plant miRNAs: first, at least 65% of nucleotides establish a base pair (including G–U pairs) with the opposite arm of the hairpin; second, no more than three consecutive nucleotides are unpaired; and third, no base pairing is allowed within the miRNA or miRNA* sequence.

Sequences with more than 80% overlap on opposing arms of the hairpin were classified as miRNA and miRNA* groups, with the more abundant sequence classified as miRNA and the other as miRNA*. If no difference in abundance was observed, the group was treated as ambiguous. Finally, we filtered for loci at which, first, at least 80% of sRNA sequences from a strand could be mapped to hairpins on the same strand following the criteria listed above, and second, the probability of obtaining the observed minimum free energy from 100 randomly shuffled sequences of the same length and dinucleotide composition (using the randfold program³⁰) was 0.05 or lower.

Identification of putative miRNA targets. To identify potential target mRNAs for each candidate miRNA we ran FASTA searches against a published *Chlamydomonas* transcripts collection (Chlre3_1.GeneCatalogTranscripts.6JUL06.fasta, available from <http://genome.jgi-psf.org>) and filtered results for compliance with the following criteria: first, no more than four mismatches (counting G–U pairs as 0.5 mismatch) between the miRNA and the target site and none at positions 10–11 (cleavage site); second, no adjacent mismatches in positions 2–12 and no more than two adjacent mismatches elsewhere; third, no more than one bulge in the target RNA and no bulge in the miRNA; and fourth, a ratio of the minimum free energy of the miRNA/target (hetero-)duplex to the minimum free energy of the perfectly paired (homo-)duplex of 0.7 or higher.

24. Wegener, D., Treier, U. & Beck, C. F. Procedures for the generation of mature *Chlamydomonas reinhardtii* zygotes for molecular and biochemical analyses. *Plant Physiol.* **90**, 512–515 (1989).
25. Chappell, L., Baulcombe, D. & Molnar, A. In *Current Protocols in Microbiology* (eds Coico, R. et al.) 16H.2.1–16H.2.17 (Wiley, Hoboken, NJ, 2005).
26. Baumberger, N. & Baulcombe, D. C. *Arabidopsis* ARGONAUTE1 is an RNA Slicer that selectively recruits micro RNAs and short interfering RNAs. *Proc. Natl Acad. Sci. USA* **102**, 11928–11933 (2005).
27. Margulies, M. et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376–380 (2005).
28. Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 1–19 (2004).
29. Hofacker, I. L. et al. Fast folding and comparison of RNA secondary structures. *Monatsh. Chem.* **125**, 167–188 (1994).
30. Bonnet, E., Wuyts, J., Rouze, P. & Van de Peer, Y. Evidence that microRNA precursors, unlike other non-coding RNAs, have lower folding free energies than random sequences. *Bioinformatics* **20**, 2911–2917 (2004).

LETTERS

A microRNA component of the p53 tumour suppressor network

Lin He^{1*}, Xingyue He^{1,2*}, Lee P. Lim³, Elisa de Stanchina^{1†}, Zhenyu Xuan¹, Yu Liang⁴, Wen Xue¹, Lars Zender¹, Jill Magnus³, Dana Ridzon⁴, Aimee L. Jackson³, Peter S. Linsley³, Caifu Chen⁴, Scott W. Lowe¹, Michele A. Cleary³ & Gregory J. Hannon¹

A global decrease in microRNA (miRNA) levels is often observed in human cancers^{1,2}, indicating that small RNAs may have an intrinsic function in tumour suppression. To identify miRNA components of tumour suppressor pathways, we compared miRNA expression profiles of wild-type and p53-deficient cells. Here we describe a family of miRNAs, miR-34a–c, whose expression reflected p53 status. Genes encoding miRNAs in the miR-34 family are direct transcriptional targets of p53, whose induction by DNA damage and oncogenic stress depends on p53 both *in vitro* and *in vivo*. Ectopic expression of miR-34 induces cell cycle arrest in both primary and tumour-derived cell lines, which is consistent with the observed ability of miR-34 to downregulate a programme of genes promoting cell cycle progression. The p53 network suppresses tumour formation through the coordinated activation of multiple transcriptional targets, and miR-34 may act in concert with other effectors to inhibit inappropriate cell proliferation.

The p53 tumour suppressor lies at a nexus of cellular pathways that sense DNA damage, cellular stress and improper mitogenic stimulation³. p53 integrates such signals and, in response, induces growth arrest, promotes apoptosis, blocks angiogenesis, or mediates DNA repair in a context-dependent manner⁴. The importance of p53 in preventing tumour formation is indicated by the presence of mutations in the p53 pathway in nearly all cancers⁵. Although p53 is most studied as a transcriptional activator, several reports have suggested that p53 represses the expression of specific genes⁶. Studies of p53-mediated repression have shown that both genes that modulate apoptotic responses and genes that promote cell cycle progression can be repressed by p53 (ref. 7).

miRNAs enforce post-transcriptional silencing through the RNA interference pathway⁸. p53-mediated induction of one or more miRNAs could therefore allow it to exert negative effects on gene expression indirectly. To explore the possibility that miRNAs might constitute part of the p53 tumour suppressor network, we examined miRNA expression profiles in wild-type and p53-deficient mouse embryonic fibroblasts (MEFs). Using the semi-quantitative reverse-transcription-polymerase chain reaction (QRT-PCR)⁹, we measured the expression of a panel of 145 mouse miRNAs in wild-type or p53^{-/-} MEFs that ectopically express various oncogenes (Fig. 1a, and Supplementary Fig. S1). miRNA expression was strongly affected by genetic alterations, because unsupervised clustering grouped MEFs according to their genotype (Supplementary Fig. S1). The expression of three miRNAs, miR-34a, miR-34b and miR-34c, was precisely correlated with p53 status (Fig. 1a). This raised the possibility that *mir-34* genes might be regulated by p53.

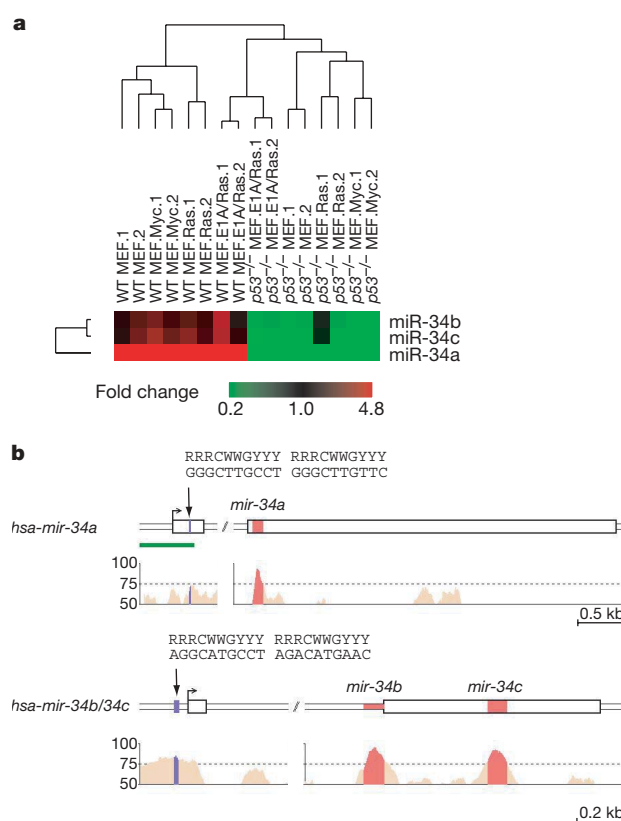


Figure 1 | Expression of miR-34 is correlated with p53 status in MEFs. **a**, An unsupervised hierarchical clustering based on miRNA expression profiles in wild-type and p53^{-/-} MEFs with the indicated additional genetic alteration. Two independently constructed cell lines (.1 and .2) were analysed in each case. The complete heat map (linear scale) is presented in Supplementary Fig. S1. **b**, Predicted gene structures for human *mir-34a* and *mir-34b/c* were generated by combining information from expressed sequence tag databases, CAGE databases and 5' rapid amplification of cDNA ends. Sequence conservation between human, mouse and rat are represented as the percentage of conservation in the Vista analysis shown in the lower panel. The promoter regions of *mir-34a* and *mir-34b/c* each contain a palindromic sequence (shown in blue) that matches the canonical p53 binding site. The green bar indicates a CpG island. kb, kilobase.

¹Watson School of Biological Sciences, Howard Hughes Medical Institute, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA. ²Program in Genetics, Stony Brook University, Stony Brook, New York 11794, USA. ³Rosetta Inpharmatics, 401 Terry Avenue N., Seattle, Washington 98109, USA. ⁴Advanced Research & Technology, Applied Biosystems, 850 Lincoln Centre Drive, Foster City, California 94404, USA. [†]Present address: Memorial Sloan-Kettering Cancer Center, 415 East 68th Street, New York, New York 10021, USA.

*These authors contributed equally to this work.

miR-34s belong to an evolutionarily conserved miRNA family, with single, recognizable orthologues in several invertebrate species (Supplementary Fig. S2). According to predicted gene structures, human miR-34a is located within exon 2 of its primary transcript, whereas miR-34b and miR-34c are located within intron 1 and exon 2, respectively, of the same primary transcript (Fig. 1b). Aside from the miRNAs themselves, the only other region of significant sequence conservation in *mir-34* genes lies in their putative promoter regions (Fig. 1b).

Because expression of miR-34s was correlated with p53 status, we asked whether miR-34s were directly regulated by p53. In MEFs expressing a tetracycline-regulated p53 short hairpin RNA (shRNA), endogenous p53 activity gradually increased over 6–8 days after repression of the shRNA⁶. Reactivation of p53 led to significant induction of both primary (pri-) *mir-34* transcripts and mature

miR-34s (Fig. 2a, and Supplementary Fig. S3A). The kinetics and magnitude of induction were comparable to those of the canonical p53 target, p21 (Supplementary Fig. S3A). Silencing of p53 in human tumour cell lines led to a roughly fourfold decrease in miR-34a levels (Supplementary Fig. S3B).

Multiple physiological stresses can induce the accumulation of p53 protein and activate p53-mediated transcriptional programmes. DNA damage mediates p53 activation mainly through post-translational modification^{10,11}. In a p53-dependent manner, both pri-miR-34s and mature miR-34s were induced by ionizing radiation in a variety of mouse tissues, including spleen, colon, thymus and kidney (Fig. 2b, Supplementary Fig. S4A, and data not shown). miR-34s were also induced after DNA damage in wild-type but not p53-null MEFs, with an amplitude and kinetics that closely resembled those of p21 (Supplementary Fig. S4B, C). Similarly, in TOV21G

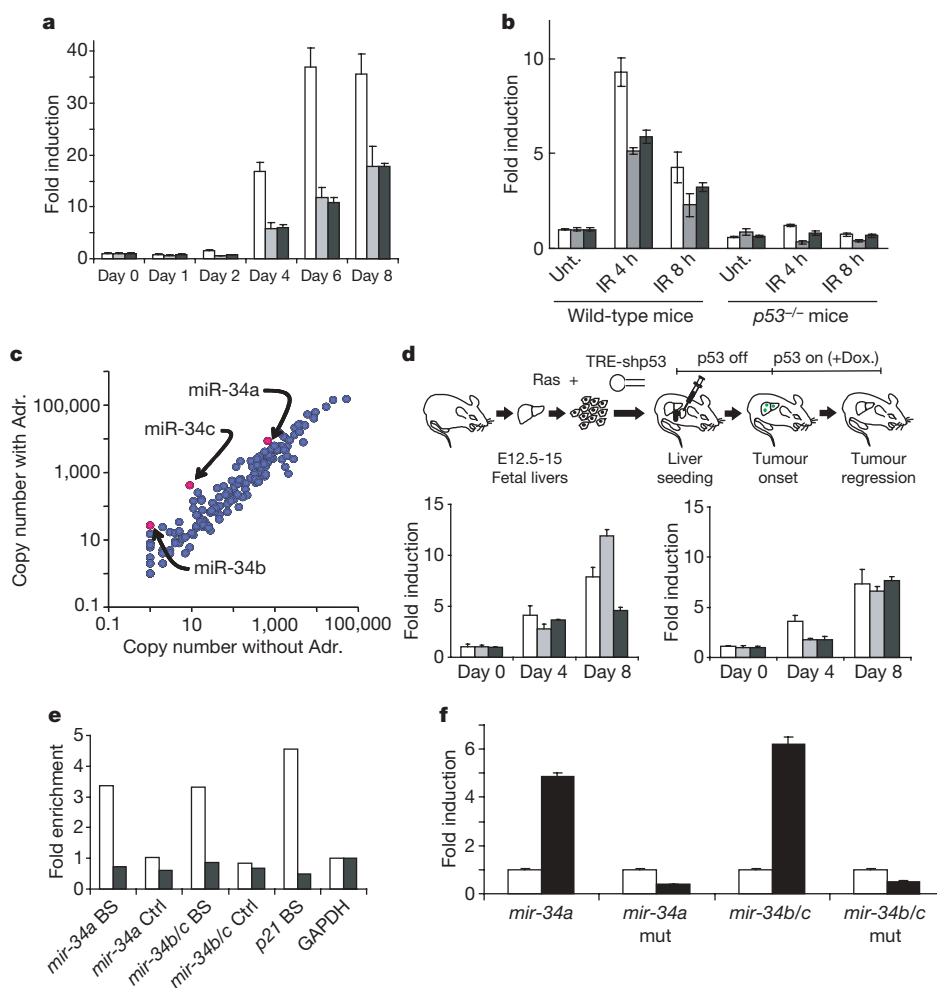


Figure 2 | Genes encoding miR-34 are direct targets of p53. **a**, miR-34 levels were measured in MEFs expressing a tetracycline-repressible p53 shRNA⁶ at the indicated times after the addition of doxycycline. White columns, mature miR-34a; grey columns, mature miR-34b; black columns, mature miR-34c. **b**, Wild-type and p53^{-/-} animals were subjected to 6 Gy of ionizing radiation (IR), and miR-34 levels (identified as in **a**) were measured in spleens by Taqman assays both before and at the indicated times after irradiation. Unt., unirradiated. **c**, A group of 191 miRNAs and selected miRNA* sequences were quantified by QRT-PCR in TOV21G cells before and after treatment with 0.1 μg ml⁻¹ adriamycin (Adr.). Results are presented in a logarithmic-scale dot plot of copy number per cell. The full data set is presented in Supplementary Table S1. **d**, Hepatocellular carcinomas were produced by combined expression of activated Ras and a conditional p53 shRNA¹³. p53 suppression was relieved by treatment with doxycycline (Dox.). Tumours were harvested at the indicated times during treatment with doxycycline, and levels of mature miR-34 were measured by

Taqman assays. Levels are plotted with respect to tumours before p53 reactivation. Left: white columns, pri-miR-34a; grey columns, pri-miR-34b/34c; black columns, mp21. Right: column colours as in **a**. **e**, ChIPs were performed with p53 antibodies on wild-type MEFs (white columns) or p53^{-/-} MEFs (black columns) treated with adriamycin. BS indicates quantification of the fragment containing the predicted p53 binding site in the *mir-34a*, *mir-34b/c* or *p21* promoter regions, and Ctrl indicates a 3' fragment from the same gene. Signals were normalized to glyceraldehyde-3-phosphate dehydrogenase (GAPDH) for each genotype. **f**, Firefly luciferase coding sequences were placed under the transcriptional control of human *mir-34a* or *mir-34b/c* promoter elements containing either wild-type or mutant (as indicated) p53 binding sites. These reporters were co-transfected with either control (white columns) or human p53 expression plasmids (black columns). Transfections were normalized by using a simultaneously delivered *Renilla* luciferase expression plasmid, pRLTK. In all cases, error bars indicate s.d. ($n = 3$).

cells, a human ovarian cancer cell line, members of the miR-34 family had one of the highest levels of induction after DNA damage among the miRNAs examined (Fig. 2c, and Supplementary Table S1).

Oncogene activation often induces the p53 pathway through induction of ARF (alternative reading frame)¹². Such a response was evident in a mouse hepatocellular carcinoma model, in which tumorigenesis driven by activated Ras required continuous suppression of p53 by an inducible shRNA¹³. Repression of the shRNA allowed Ras-mediated activation of endogenous p53 and resulted in the senescence of tumour cells¹³. Under these circumstances, all three miR-34s were strongly induced, supporting their regulation by p53 *in vivo* (Fig. 2d). Similarly, oncogenic stress can induce miR-34a in cultured primary human fibroblasts (data not shown).

One of the few conserved regions within the genes encoding miR-34 family members contains a match to the canonical p53 binding site (see Fig. 1b). To test p53 binding to these sites, we performed chromatin immunoprecipitation (ChIP). In wild-type MEFs, in which p53 activity was induced by DNA damage, regions of the genes encoding both miR-34a and miR-34b/c that contained putative p53 binding sites were enriched in p53 immunoprecipitates. This enrichment was absent from similarly treated p53-null MEFs (Fig. 2e). These same sites were previously detected in genome-wide chromatin occupancy experiments with p53, though their significance for regulation of miR-34s was not noted¹⁴. To examine the potential of *mir-34* promoters to confer p53 regulation, we inserted fragments of the mouse or human genes encoding miR-34a or miR-34b/c upstream of a luciferase-coding region. Co-transfection of these reporters with a p53 expression vector robustly stimulated luciferase expression (Fig. 2f, and data not shown). Mutation of p53 binding sites in these reporters negated this induction. These findings indicate that miR-34a and miR-34b/c are direct transcriptional targets of p53.

Two major endpoints of p53 activation are apoptosis and growth arrest (either cell cycle arrest or senescence). The ectopic expression of either *mir-34a* or *mir-34b/c* in IMR90 cells led to substantial inhibition of growth (Fig. 3a). This was attributable to effects on cell proliferation: the fraction of S-phase cells decreased, and the fraction

of cells in G1 and G2 increased (Fig. 3b). We also noted distinctive morphological alterations characteristic of cellular senescence (Fig. 3c), and about 60% of infected cells stained positively for a senescence marker, SA- β -Gal, at 6 days after selection (Fig. 3d). Importantly, all of these effects were seen with *mir-34* expression levels similar to those achieved after p53-mediated induction (Supplementary Fig. S5). Transfection of miR-34 miRNAs, but not that of miR-34s containing seed mutations, also led to G1 arrest in immortalized mouse cells and in human tumour cell lines including NIH-3T3, HCT-116, A549 and TOV21G (Supplementary Fig. S6, data not shown). These studies indicate that arrest can be induced independently of the integrity of major tumour suppressor pathways, at least in some cell lines (Fig. 3, and data not shown). Ectopic delivery of miR-34a also sensitized MEFs to apoptosis in response to genotoxic stress, although the effects were not as pronounced as the growth arrest induced by this miRNA (data not shown).

miRNAs often decrease the mRNA levels of direct regulatory targets¹⁵. After transfection of miR-34a, b or c into a panel of four tumour cell lines, a cluster of genes was specifically downregulated at 24 h after transfection (Supplementary Fig. S7 and Supplementary Table S2), with some genes showing significant repression as early as 10 h after transfection. These genes were highly enriched for transcripts with 3' untranslated regions (UTRs) containing complements to miR-34 seed hexamers. On the basis of functional annotation, genes involved in control of the cell cycle were strongly overrepresented among this set ($P < 3.4 \times 10^{-20}$; Supplementary Fig. S7). A selection of candidate targets, including cyclin E2 (CCNE2), cyclin-dependent kinase 4 (CDK4) and the hepatocyte growth factor receptor (MET)¹⁶, were validated by western blotting (Fig. 4a, and Supplementary Fig. S8). On transfection of miR-34a into A549 and HCT116 cells, we observed the expected twofold to fourfold decrease in each target examined¹⁵. To test whether regulation was direct, we fused the 3' UTRs of these selected targets to luciferase. Co-transfection with miR-34a but not miR-124a specifically decreased luciferase levels from each reporter (Fig. 4b). Mutations in seed complementary sites fully rescued repression for both CDK4 and MET. For cyclin E2, mutation of the single best seed complementary site had only a partial

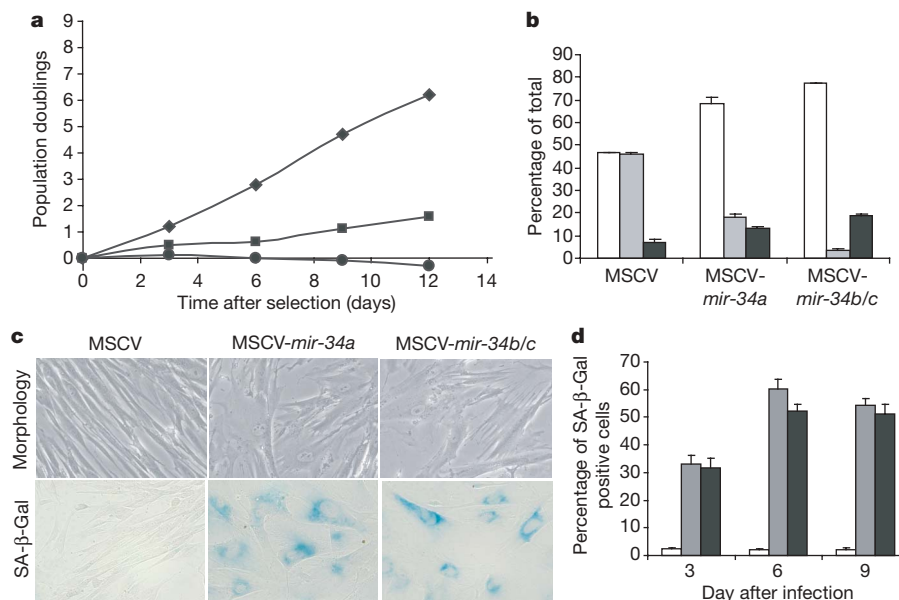


Figure 3 | miR-34 family miRNAs mediate growth arrest in a variety of cell types. **a**, Proliferation of IMR90 cells was measured as cumulative population doublings after retroviral delivery of vectors directing the expression of primary miR-34a (squares), miR-34b/c (circles) or a control MSCV vector (diamonds). Measurements were initiated immediately after selection with puromycin. **b**, Cell cycle analysis was performed 1 day after selection with puromycin by BrdU/FACS on IMR90 cells engineered as in

a. White columns, G1; grey columns, S; black columns, G2/M. **c**, IMR90 cells engineered to express pri-miR-34a or pri-miR-34b/c showed morphological alterations similar to those seen in senescent cells. **d**, Percentages of SA- β -Gal-positive cells were determined at 3, 6 and 9 days after the completion of selection with puromycin. White columns, MSCV; grey columns, MSCV-mir-34a; black columns, MSCV-mir-34b/c. In all cases, error bars indicate s.e.m. ($n = 3$).

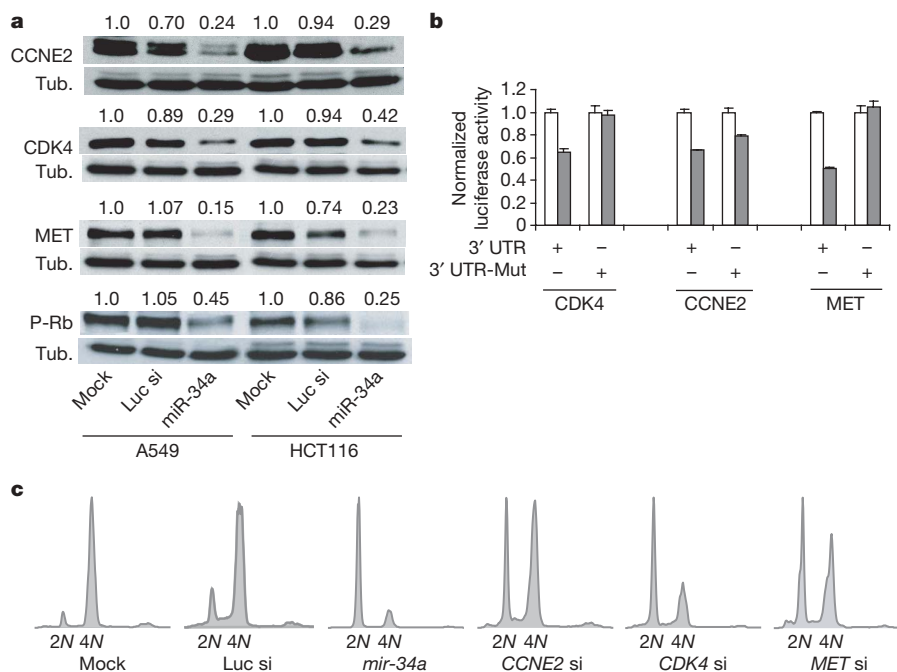


Figure 4 | miR-34 regulates a programme of cell cycle and DNA damage response genes. **a**, Western blots were used to measure protein levels after miR-34 delivery for multiple candidate targets identified in the cell cycle overlapping gene set in Supplementary Fig. S6. Tub., tubulin. **b**, Reporter plasmids in which the luciferase coding sequence had been fused to the 3' UTR of CDK4, CCNE2 or MET, as indicated, were transfected into HeLa cells in conjunction with either miR-34a (grey columns) or miR-124a (white columns) siRNAs. Luciferase activity was normalized relative to a

simultaneously transfected *Renilla* expression plasmid. In each case 3' UTR-Mut indicates the introduction of alterations into the seed complementary sites shown in Supplementary Fig. S8. Error bars indicate s.e.m. ($n = 3$). **c**, HCT116 Dicer^{ex5} cells were transfected with siRNAs targeting CDK4, CCNE2 and MET, and cell cycle effects were analysed as described in Supplementary Fig. S6. The somewhat less efficacious arrest on transfection with CCNE2 siRNA could reflect a partly redundant function or less potent suppression of its mRNA.

effect (Fig. 4b), indicating either the presence of other relevant seed complements or a combination of both direct and indirect effects of miR-34. Silencing of these selected miR-34 targets by using siRNAs led to a substantial arrest in G1 (Fig. 4c), partly phenocopying activation of their upstream regulator. Ectopic miR-34 delivery caused a decrease in levels of phosphorylated retinoblastoma gene product (Rb), consistent with lowered activity of both CDK4 and CCNE2 complexes (Fig. 4a). Repression of CDK4 and CCNE2 has previously been noted after p53 activation in PC3 cells⁷. Our results indicate the possibility that p53 might repress these genes indirectly by the induction of miR-34.

We also noted a significant overlap between miR-34-regulated genes and those whose expression is altered after DNA damage (Supplementary Fig. S7). This was seen both for genes that increased after either miR-34 delivery or DNA damage ($P < 7 \times 10^{-65}$) and for those that decreased in response to either treatment ($P < 1.8 \times 10^{-73}$). Although strong seed enrichment was seen in the mutually downregulated set, seed enrichment was not seen in the mutually upregulated set, indicating that such increases in expression might be secondary effects of miR-34.

Activation of p53 leads to the coordinated induction of multiple downstream effectors, many of which act in a partly or fully redundant manner. A classic example is p53-dependent apoptosis, which depends not only on the induction of *bax*¹⁷ but also on *puma* and *noxa*¹⁸. For p53-mediated growth arrest, induction of the CDK inhibitor p21 is clearly important. However, p21 loss does not completely negate the ability of p53 to halt proliferation^{19,20}. This demonstrates the existence of redundant or cooperating pathways that contribute to p53-mediated arrest in G1. Recent studies have identified several p53 targets, including Gadd45a (for growth arrest and DNA-damage-inducible), 14-3-3 and Reprimo, which have been proposed to collaborate with p21, primarily to trigger arrest in G2 in specific cell types³. At least in some contexts, miR-34s can exert their growth inhibitory effects in the absence of p21, because HCT-116 cells

lacking p21 are susceptible to miR-34 (Supplementary Fig. S6B). This is especially important because reports have implicated p21 in p53-mediated repression²¹.

Our data identify the miR-34 family of miRNAs as direct targets of p53 that possess anti-proliferative potential. It is likely that miR-34s mediate this response through additive or synergistic effects of multiple targets, because many components of the cell cycle machinery are affected after the manipulation of miR-34 levels. The effects of miR-34s may also extend to the other arm of the p53 response, given a recent report and our findings (data not shown) that miR-34a can enhance apoptotic responses in some cell types²². Thus, the actual phenotypic output of miR-34 activation may vary by cell type depending on the spectrum of its targets that are available for repression.

In accord with their regulation by p53, comparatively low levels of miR-34s are observed in human tumours and cancer cell lines, which have a high frequency of functional p53 deficiency. Although selective pressures for miR-34 deletion in human cancers may be alleviated by frequent p53 mutations, deletion of miRNAs of the miR-34 family has been reported in several human tumours and cancer cell lines^{22,23}. In fact, the human gene encoding miR-34a maps to 1p36, a locus frequently deleted in human cancers. Recently, one gene within this locus, *CHD5*, has been implicated in its tumour suppressive activity²⁴. *CHD5* has been proposed to act upstream of p53 by regulating its expression in response to various p53-inducing stimuli. Deletions at 1p36 are often quite large and can encompass both *CHD5* and the *mir-34a* locus, as well as other genes. Thus, 1p36 genomic lesions might affect the p53 pathway at multiple levels, both upstream and downstream of p53 activation.

Although dozens of p53 targets have been identified in mammals, very few are evolutionarily conserved in *Drosophila* and *Caenorhabditis elegans*, both of which retain homologues of the p53 pathway²⁵. miR-34 is one of only 18 mammalian miRNA families²⁶ that are also present in flies and worms. This raises the

possibility that the link between p53 and this non-coding RNA target may have arisen early in the evolution of the p53 network and may be important in p53 function in diverse species.

METHODS SUMMARY

miRNA expression analyses. TaqMan MicroRNA assays were used to quantify levels of mature miRNAs as described previously⁹. Primary miR-34 transcript levels were determined with the SuperScript III SYBR Green One-Step qRT-PCR system (Invitrogen).

Proliferation and cell cycle analysis. Primary IMR90 cells were infected with murine stem cell virus (MSCV) retroviral vectors expressing *mir-34a* or *mir-34b/c* and were used for a variety of growth assays. Cell cycle distributions were measured with the bromodeoxyuridine (BrdU) APC flow kit (BD Biosciences). Tumour-derived cell lines, including A549 and HCT116, were transiently transfected with luciferase siRNA GL-3 or miR-34a-c duplexes (100 nM). After 24 h, transfected populations were treated for 16–20 h with nocodazole (100 ng ml⁻¹). Cell cycle profiles were determined by staining with propidium iodide, followed by fluorescence-activated cell sorting (FACS).

ChIP and promoter reporter assays. For reporter assays, putative promoter regions of human *mir-34a* and *mir-34b/c* were amplified from genomic DNA and cloned upstream of firefly luciferase in the pGL4 plasmid (Promega). Mutagenesis of p53 binding sites was performed with the QuickchangeXL mutagenesis kit (Stratagene). pGL4 vectors containing wild-type and mutant *mir-34* promoters were transfected into the HCT116 p53^{-/-} cells, together with either the pLPC-p53 vector or an empty vector. The promoter activity of each construct was determined by dual luciferase assay (Promega) 24 h after transfection.

Microarray profiling and western analysis. miRNA duplexes (100 nM) were transfected into all cell lines. RNA was isolated 6–24 h after transfection, and microarray analysis was performed as described¹⁵. Antibodies against CDK4 (Santa Cruz), CCNE2, MET (Cell Signaling) and phospho-Rb (Santa Cruz) were used in western analysis in accordance with the manufacturer's instructions. Anti-tubulin antibodies (Sigma) were used to test for equal loading.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 17 March; accepted 17 May 2007.

Published online 6 June 2007.

1. Lu, J. *et al.* MicroRNA expression profiles classify human cancers. *Nature* **435**, 834–838 (2005).
2. Thomson, J. M. *et al.* Extensive post-transcriptional regulation of microRNAs and its implications for cancer. *Genes Dev.* **20**, 2202–2207 (2006).
3. Levine, A. J., Hu, W. & Feng, Z. The p53 pathway: what questions remain to be explored? *Cell Death Differ.* **13**, 1027–1036 (2006).
4. Ko, L. J. & Prives, C. p53: puzzle and paradigm. *Genes Dev.* **10**, 1054–1072 (1996).
5. Hollstein, M., Sidransky, D., Vogelstein, B. & Harris, C. C. p53 mutations in human cancers. *Science* **253**, 49–53 (1991).
6. Dickins, R. A. *et al.* Probing tumor phenotypes using stable and regulated synthetic microRNA precursors. *Nature Genet.* **37**, 1289–1295 (2005).
7. Spurgers, K. B. *et al.* Identification of cell cycle regulatory genes as principal targets of p53-mediated transcriptional repression. *J. Biol. Chem.* **281**, 25134–25142 (2006).
8. Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281–297 (2004).

9. Chen, C. *et al.* Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Res.* **33**, e179 (2005).
10. Fei, P. & El-Deiry, W. S. P53 and radiation responses. *Oncogene* **22**, 5774–5783 (2003).
11. Giaccia, A. J. & Kastan, M. B. The complexity of p53 modulation: emerging patterns from divergent signals. *Genes Dev.* **12**, 2973–2983 (1998).
12. Sherr, C. J. & Weber, J. D. The ARF/p53 pathway. *Curr. Opin. Genet. Dev.* **10**, 94–99 (2000).
13. Xue, W. *et al.* Senescence and tumour clearance is triggered by p53 restoration in murine liver carcinomas. *Nature* **445**, 656–660 (2007).
14. Wei, C. L. *et al.* A global map of p53 transcription-factor binding sites in the human genome. *Cell* **124**, 207–219 (2006).
15. Lim, L. P. *et al.* Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* **433**, 769–773 (2005).
16. Lewis, B. P., Burge, C. B. & Bartel, D. P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15–20 (2005).
17. Miyashita, T. & Reed, J. C. Tumor suppressor p53 is a direct transcriptional activator of the human *bax* gene. *Cell* **80**, 293–299 (1995).
18. Villunger, A. *et al.* p53- and drug-induced apoptotic responses mediated by BH3-only proteins puma and noxa. *Science* **302**, 1036–1038 (2003).
19. Brugarolas, J. *et al.* Radiation-induced cell cycle arrest compromised by p21 deficiency. *Nature* **377**, 552–557 (1995).
20. Deng, C., Zhang, P., Harper, J. W., Elledge, S. J. & Leder, P. Mice lacking p21CIP1/WAF1 undergo normal development, but are defective in G1 checkpoint control. *Cell* **82**, 675–684 (1995).
21. Lohr, K., Moritz, C., Contente, A. & Dobbelstein, M. p21/CDKN1A mediates negative regulation of transcription by p53. *J. Biol. Chem.* **278**, 32507–32516 (2003).
22. Welch, C., Chen, Y. & Stallings, R. L. MicroRNA-34a functions as a potential tumor suppressor by inducing apoptosis in neuroblastoma cells. *Oncogene* advance online publication, doi:10.1038/sj.onc.1210293 (12 February 2007).
23. Calin, G. A. *et al.* MicroRNA profiling reveals distinct signatures in B cell chronic lymphocytic leukemias. *Proc. Natl Acad. Sci. USA* **101**, 11755–11760 (2004).
24. Bagchi, A. *et al.* CHD5 is a tumor suppressor at human 1p36. *Cell* **128**, 459–475 (2007).
25. Sutcliffe, J. E. & Brehm, A. Of flies and men; p53, a tumour suppressor. *FEBS Lett.* **567**, 86–91 (2004).
26. Ruby, J. G. *et al.* Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. *Cell* **127**, 1193–1207 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank members of the Hannon and Lowe laboratories and the Rosetta Biology group for helpful input; M. Zhang and J. Burchard for bioinformatic analysis; J. Guo, C. Raymond and K. Niemeyer for miRNA quantification; J. Schelter and M. Kibukawa for cell cycle analyses and gene expression profiling; R. Diaz, M. Mehaffey, F. Huynh and the Rosetta Gene Expression Laboratory for technical assistance; and R. Dickins, J. Kurland, M. McCurrach, K. Diggins, A. Chicas, B. Stillman and B. Vogelstein for providing reagents and protocols. L.H. is a Fellow of the Helen Hay Whitney Foundation and is supported by a K99 grant from the NCI. S.W.L. and G.J.H. are supported by a program project grant from the NCI and are investigators of the Howard Hughes Medical Institute. This work was also supported in part by a gift from K. W. Davis.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.A.C. (michele_clearly@merck.com) or G.J.H. (hannon@cshl.org).

METHODS

Quantification of miRNAs with real-time PCR. TaqMan MicroRNA assays were used to quantify the level of mature miRNAs as described previously^{3,27}. In miRNA profiling experiments, each reverse transcriptase (RT) reaction contained 3.75 ng of purified total RNA, 50 nM stem-loop RT primer, 1 × RT buffer, dNTPs (each at 0.25 mM), 3.33 U μl⁻¹ MultiScribe reverse transcriptase and 0.25 U μl⁻¹ RNase inhibitor (Applied Biosystems). The reactions were incubated for 30 min at 16 °C, 30 min at 42 °C, and 5 min at 85 °C. Real-time PCR reactions for each miRNA (10 μl volume) were performed in quadruplicate, and each 10-μl reaction mixture included 2 μl of diluted RT product (1:2 dilution), 5 μl of 2 × TaqMan Universal PCR Master Mix, 0.2 μM TaqMan probe, 1.5 μM forward primer, and 0.7 μM reverse primer, respectively (Applied Biosystems). Reactions were incubated in an Applied Biosystems 7900HT Fast Real-Time PCR system in 384-well plates at 95 °C for 10 min, followed by 40 cycles at 95 °C for 15 s and 60 °C for 1 min. The threshold cycle (*C_t*) is defined as the fractional cycle number at which the fluorescence exceeds the fixed threshold of 0.2. Primary transcripts level of *mir-34a* and *mir-34b/c* were determined with the SuperScript III SYBR Green One-Step qRT-PCR system (Invitrogen). Primers that amplify the *mir-34a* pri-miRNA, *mir-34b/c* pri-miRNA and the control *β-actin* mRNA were designed with Primer Express software, v. 2: *hsa-pri-mir-34a* forward primer, 5'-CCTCCAAGCCAGCTCAGTTG-3'; *hsa-pri-mir-34a* reverse primer, 5'-TGACTTTGGTCCAATTCCTGTTG-3'; *hsa-pri-mir-34b/c* forward primer, 5'-GCTCTTTGTCCCTCCTGCTAGA-3'; *hsa-pri-mir-34b/c* reverse primer, 5'-GTGGGCGGTCCCTGAAG-3'; *mmu-pri-mir-34a* forward primer, 5'-CTGTGCCCTCTTGCAAAAGG-3'; *mmu-pri-mir-34a* reverse primer, 5'-GGACATTCAGGTGAGGGTCTTG-3'; *mmu-pri-mir-34b/c* forward primer, 5'-GGCAGGAAGGCTCCAGATG-3'; *mmu-pri-mir-34b/c* reverse primer, 5'-CCTCACTGTTTCATATGCCCATTC-3'. The ratios of RNA species in each sample were determined in triplicate with the use of an ABI 7900HT TaqMan sequence detector following the standard curve method.

Hierarchical clustering for miRNA expression profiling. Expression of 145 mouse miRNAs was characterized and the data were subjected to a series of adjustment and filtering before hierarchical clustering. Assays with *C_t* values greater than 35 were treated as 35, and low-expressing miRNAs were then removed from the analysis if their average *C_t* values across the samples were between 34 and 35. Four miRNAs (miR-30d, miR-148b, miR-320 and let-7d) that were least variable among the 16 samples used in this study were selected as internal references, and ΔC_t between the *C_t* of each miRNA and the average of these four references for each sample were calculated. Data from each of the resulting 115 miRNAs were median-centred, and both miRNAs and samples were clustered by using the average linkage method under the correlation similarity metric.

Chromatin immunoprecipitation. Wild-type MEF and *p53*^{-/-} MEF cultures were grown to 70–80% confluence and then treated with doxorubicin (0.5 μg ml⁻¹) for 24 h. After being washed with PBS, cells were crosslinked with 1% formaldehyde for 15 min at room temperature. Crosslinking was stopped by the addition of glycine to 125 mM final concentration. Cells were washed twice with cold PBS and then harvested in lysis buffer first (20 mM Tris-HCl pH 8.0, 85 mM KCl, 1 mM EDTA, 0.5 mM EGTA, 0.5% Nonidet P40, and protease inhibitors), and the pellets were then dissolved in nuclear lysis buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 1% SDS, and protease inhibitors). Samples were sonicated to generate DNA fragments less than 500 base pairs in length. Before immunoprecipitation, nuclear extracts were precleared at 4 °C with 50 μl of 50% Protein A-Sepharose slurry for 30 min. p53 antibody (CM5; Novocastra) was then added to form complexes with p53 protein and associated chromatin. These immunocomplexes were recovered with Protein A-Sepharose beads (Upstate), and the associated DNA was purified by extraction with phenol/chloroform. The enrichment across the putative p53 binding sites at both *mir-34a* and *mir-34b/c* were then tested by real-time PCR analysis. Primers 5'-CAGCTGGAGGAGGATCGA-3' and 5'-TCCCAAGCCCCCAATCT-3' were used to amplify the *mir-34a* promoter regions containing the putative p53 binding sites; primers 5'-GTTGATCCTGCCACAGTTACTAGA-3' and 5'-ATTAAACATGAGTCTCCCTGGTCTCT-3' were used to amplify the *mir-34b/c* promoter regions containing the putative p53 binding sites. Two sets of primers designed to amplify the 3' end of the gene encoding the *mir-34a* precursor and the gene encoding the *mir-34b/c* precursor, respectively, were used as controls.

Proliferation and cell cycle analysis. Primary IMR90 cells and MEFs were infected with MSCV retroviral vectors directing the expression of *mir-34a* or *mir-34b/c* from the long terminal repeat promoter. Infected cells were selected by puromycin for 2 days, left to recover for 12 h and then used for a variety of growth assays. Growth curves were measured by determining accumulative population doublings over a course of 12 days after puromycin selection. At day 1 after selection, IMR90 cells were labelled with BrdU for 6 h, and G1, S and

G2/M populations were measured by the BrdU APC flow kit (BD Biosciences). SA-β-Gal staining was conducted for IMR90 cells at 3, 6 and 9 days after selection with puromycin. Cells were fixed with 0.5% glutaraldehyde in PBS for 15 min at room temperature. After being washed with PBS/MgCl₂ (1 mM MgCl₂, pH 6.0), cells were stained in X-Gal (1 mg ml⁻¹ X-Gal, 0.1 M K₃Fe(CN)₆, 0.1 M K₃Fe(CN)₆·3H₂O, 1 mM MgCl₂, in PBS, pH 6.0). Tumour-derived cell lines including A549, HCT116 and HCT116 *p21*^{-/-} cells were transfected with luciferase siRNA GL-3 or *mir-34a-c* siRNA duplexes (100 nM). At 24 h after transfection, the cells were treated with nocodazole (100 ng ml⁻¹) for 16–20 h. Cell cycle distributions were measured by staining with propidium iodide, followed by FACS.

Microarray profiling. Cells were plated 24 h before transfection. HCT116 Dicer^{Ex5} cells were transfected in six-well plates with duplexed RNAs with the use of Lipofectamine 2000 (Invitrogen). DLD-1, HeLa, TOV21G and A549 cells were transfected by using SilentFect (Bio-Rad). Duplexes were used at final concentrations of 100 nM for all cell lines. RNA was isolated 6–24 h after transfection, and microarray analysis was performed as described²⁸.

Gene set analysis methods. miRNA-regulated transcripts were identified in microarray gene expression signatures using a *P*-value cut-off (*P* < 0.01). miRNA downregulated transcripts were defined by the intersection of downregulated transcripts in all the lines tested. Downregulated transcripts were tested for enrichment relative to a background set with the use of the hypergeometric distribution. miRNA target regulation was measured by enrichment of transcripts containing miRNA hexamer seed strings (stretches of six contiguous bases complementary to miRNA seed region nucleotides 1–6, 2–7 or 3–8) in transcripts having annotated 3' UTRs. Biological function was categorized by enrichment of transcripts from Gene Ontology Biological Process functional categories (<http://www.geneontology.org/>). The set of genes on the microarray was used as a background set.

Cell lines. Wild-type MEFs and *p53*^{-/-} MEFs were isolated from embryonic day (E)13.5 embryos. HCT116, HCT116 *p53*^{-/-}, HCT116 Dicer^{Ex5}, HCT116 *p21*^{-/-} and DLD-1 Dicer^{Ex5} cells were provided by B. Vogelstein's group. HEPG2, A549, 3T3 and TOV21G cells were acquired from ATCC. HEPG2, A549 and TOV21G cells were engineered to contain an empty vector or a vector expressing a hairpin targeting human p53. For these cells, the p53 shRNA was a 19-mer stem-loop cloned under the control of a human H1 promoter²⁹ and shuttled into the pLenti6 Block-It vector (Invitrogen). The engineered lines were generated by stable transduction with packaged lentiviral particles. The knockdown level of p53 was about 95%.

Western analysis. Antibodies against CDK4 (c-22; dilution 1:1,000; Santa Cruz), CCNE2 (dilution 1:1,000; Cell Signaling), MET (25H2; dilution 1:1,000; Cell Signaling) and phospho-Rb (dilution 1:2000; Santa Cruz) were used in western analysis in accordance with the manufacturer's instruction. Tubulin (Sigma) was blotted for a loading control, as well as for subsequent quantification.

miR-34-mediated suppression reporter assays. The 3' UTRs from human CDK4, CCNE2 and MET were amplified from human genomic DNA (Promega) and individually cloned into pEntr/D (Invitrogen) by directional TOPO cloning. Seed regions were mutated to remove all complementarity to nucleotides 1–7 of miR-34s by using the QuickchangeXL Mutagenesis Kit (Stratagene). Both wild-type and mutant 3' UTRs were transferred into pGL3-TK (Promega) with the use of Gateway cloning (Invitrogen). HeLa cells were co-transfected with reporter constructs and miRNAs (*mir-34a* or *mir-124a*) in the form of siRNAs using LT1 and TKO (Mirus). pRLTK (Promega) was also transfected as a normalization control. Cells were lysed 24 h after transfection, and ratios between firefly luciferase and *Renilla* luciferase activity were measured with a dual luciferase assay (Promega).

Promoter reporter assays. The putative promoter regions of human *mir-34a* and *mir-34b/c* were amplified from genomic DNA and cloned into pGL4 vector (Promega). Mutagenesis of p53 binding sites was performed with the QuickchangeXL mutagenesis kit (Stratagene). HCT116 *p53*^{-/-} cells were transfected with the heterologous reporter vectors and pRL-TK along with either the pLPC-p53 vector or pLPC control vector. Ratios of firefly luciferase to *Renilla* luciferase activity were determined with a dual luciferase assay (Promega) 24 h after transfection.

27. Raymond, C. K., Roberts, B. S., Garrett-Engle, P., Lim, L. P. & Johnson, J. M. Simple, quantitative primer-extension PCR assay for direct monitoring of microRNAs and short-interfering RNAs. *RNA* **11**, 1737–1744 (2005).
28. Jackson, A. L. *et al.* Expression profiling reveals off-target gene regulation by RNAi. *Nature Biotechnol.* **21**, 635–637 (2003).
29. Brummelkamp, T. R., Bernards, R. & Agami, R. A system for stable expression of short interfering RNAs in mammalian cells. *Science* **296**, 550–553 (2002).

Dual E1 activation systems for ubiquitin differentially regulate E2 enzyme charging

Jianping Jin¹, Xue Li², Steven P. Gygi² & J. Wade Harper¹

Modification of proteins with ubiquitin or ubiquitin-like proteins (UBLs) by means of an E1–E2–E3 cascade controls many signalling networks^{1–3}. Ubiquitin conjugation involves adenylation and thioesterification of the carboxy-terminal carboxylate of ubiquitin by the E1-activating enzyme Ube1 (Uba1 in yeast), followed by ubiquitin transfer to an E2-conjugating enzyme through a trans-thiolation reaction^{4–7}. Charged E2s function with E3s to ubiquitinate substrates¹. It is currently thought that Ube1/Uba1 is the sole E1 for charging of E2s with ubiquitin in animals and fungi^{1,8}. Here we identify a divergent E1 in vertebrates and sea urchin, Uba6, which specifically activates ubiquitin but not other UBLs *in vitro* and *in vivo*. Human Uba6 and Ube1 have distinct preferences for E2 charging *in vitro*, and their specificity depends in part on their C-terminal ubiquitin-fold domains, which recruit E2s. In tissue culture cells, Uba6 is required for charging a previously uncharacterized Uba6-specific E2 (Use1), whereas Ube1 is required for charging the cell-cycle E2s Cdc34A and Cdc34B. Our data reveal unexpected complexity in the pathways that control the conjugation of ubiquitin, in which dual E1s orchestrate the charging of distinct cohorts of E2s.

It is currently believed that single E1 enzymes within each species function to initiate the activation and conjugation of individual UBLs including ubiquitin, Nedd8 and small ubiquitin modifier (SUMO). In the case of ubiquitin itself, classic work using rodent ts20 and ts85 cells harbouring temperature-sensitive mutations in the X-chromosome-linked *UBE1* gene demonstrated loss of ubiquitin conjugation, decreased protein turnover, and cell cycle arrest when shifted to non-permissive temperature^{9,10}. Similarly, in budding yeast, a single ubiquitin E1 protein, Uba1p, is required for sporulation and cell proliferation¹¹. Mice express a second testes-specific E1, Ubely, which is about 90% identical to Ube1x (Fig. 1a, and Supplementary Fig. S1a) and is thought to act redundantly with Ube1x during spermatogenesis¹². However, humans, primates and zebrafish contain only a single *UBE1* gene (Fig. 1a, and Supplementary Fig. S1a). These and other data have led to the presumption that a single activating system exists for the charging of E2s with ubiquitin (see Supplementary Information)^{1,8}.

E1s contain three domains: first, an adenylation domain composed of two ThiF-homology motifs^{13–17}, which bind ATP and the appropriate UBL; second, the catalytic cysteine domain (CCD), which is the acyl carrier for ubiquitin; and third, the C-terminal ubiquitin-fold domain (Ufd, also referred to as a β -grasp fold), which recruits specific E2s^{16,18–20}. Using the ThiF-homology motif as a search tool, we identified genes for all characterized E1 enzymes in the human genome as well as an uncharacterized protein, Uba6 (Gene ID 55236), which also contained a CCD and Ufd (Fig. 1a, and Supplementary Fig. S1a–c). Uba6 is distantly related to Ube1 (about 40% identity; Supplementary Fig. S1a, c, d). Uba6 orthologues

were found in vertebrates and the echinoderm sea urchin but were not detected in insects, worms, fungi or plants (Fig. 1a, and Supplementary Fig. S1a). Like Ube1, Uba6 messenger RNA and protein are widely expressed in human tissues and cell types (Fig. 2a, and Supplementary Fig. S2).

We tested whether recombinant Uba6 (Fig. 1b) could activate glutathione S-transferase (GST) fusions of several human UBLs, including Urm1, Fat-10 and Fub1 whose E1s are unknown. Uba6 activated GST-ubiquitin but not other GST-UBLs tested (Fig. 1c, and Supplementary Fig. S3; thioesters are indicated by a '~' bond) in an ATP- and active-site Cys 625-dependent manner, and was sensitive to reducing agent (Fig. 1d). In contrast, the closely related E1, Ube1L (Fig. 1a), was unable to activate ubiquitin (Supplementary Fig. S3a). Uba6 and Ube1 had comparable rates of ubiquitin activation (Fig. 1e, and Supplementary Fig. S3d, e) and were retrieved from cell extracts by ubiquitin-agarose^{4,6,7,21} with similar efficiencies (Fig. 1f). Thus, the *in vitro* ubiquitin activation properties of Uba6 and Ube1 are similar.

To examine activation by Uba6 *in vivo*, we took advantage of the fact that oxy-esters with UBLs are significantly more stable than thioesters at pH 7.5, allowing their purification²². Wild-type, C625S and C625A Uba6 tagged with Flag and haemagglutinin (HA) expressed at endogenous levels in 293T cells migrated more slowly than the endogenous protein (Fig. 2b), which was consistent with the Flag-HA epitope tag. However, Uba6^{C625S} produced an additional more slowly migrating band that reacted with anti-ubiquitin (but not anti-Nedd8; Fig. 2b, lane 3; Supplementary Fig. S4a). Mass spectral analysis of the more slowly migrating Flag-HA-Uba6^{C625S} protein (35% coverage) identified four ubiquitin peptides (60% coverage), but peptides from other UBLs were absent (Fig. 2e, and Supplementary Fig. S4b, c; see Methods). No ubiquitin or UBL peptides were identified in the faster-migrating Flag-HA-Uba6^{C625S} protein (31% coverage; Supplementary Fig. S4c).

We found that ubiquitin-activated Flag-HA-Uba6 could be detected and isolated from lysates at pH 4.5, which stabilizes thioesters (Fig. 2c). Formation of the more slowly migrating Flag-HA-Uba6~ubiquitin adduct required Cys 625 (lane 2) and was sensitive to reducing agent (lane 4). Moreover, the more slowly migrating form of Flag-HA-Uba6 reacted with anti-ubiquitin (lane 3). We recovered three ubiquitin peptides by mass spectrometry (50% coverage) specifically in the more slowly migrating Uba6-containing band (41% coverage, 36 peptides; Fig. 2d, e), but peptides from other UBLs were not found. Thus, Uba6 functions specifically to activate ubiquitin in human tissue culture cells.

Next we examined E1 specificity in E2 charging of a panel of 29 E2s (Fig. 3a, and Supplementary Table S1) with the use of lysine-free ubiquitin in a non-reducing gel end-point assay (see Methods). Some E2s (for example E2S) also underwent auto-ubiquitination

¹Department of Pathology, Harvard Medical School, 77 Avenue Louis Pasteur Boston, Massachusetts 02115, USA. ²Department of Cell Biology, Harvard Medical School, 200 Longwood Drive Boston, Massachusetts 02115, USA.

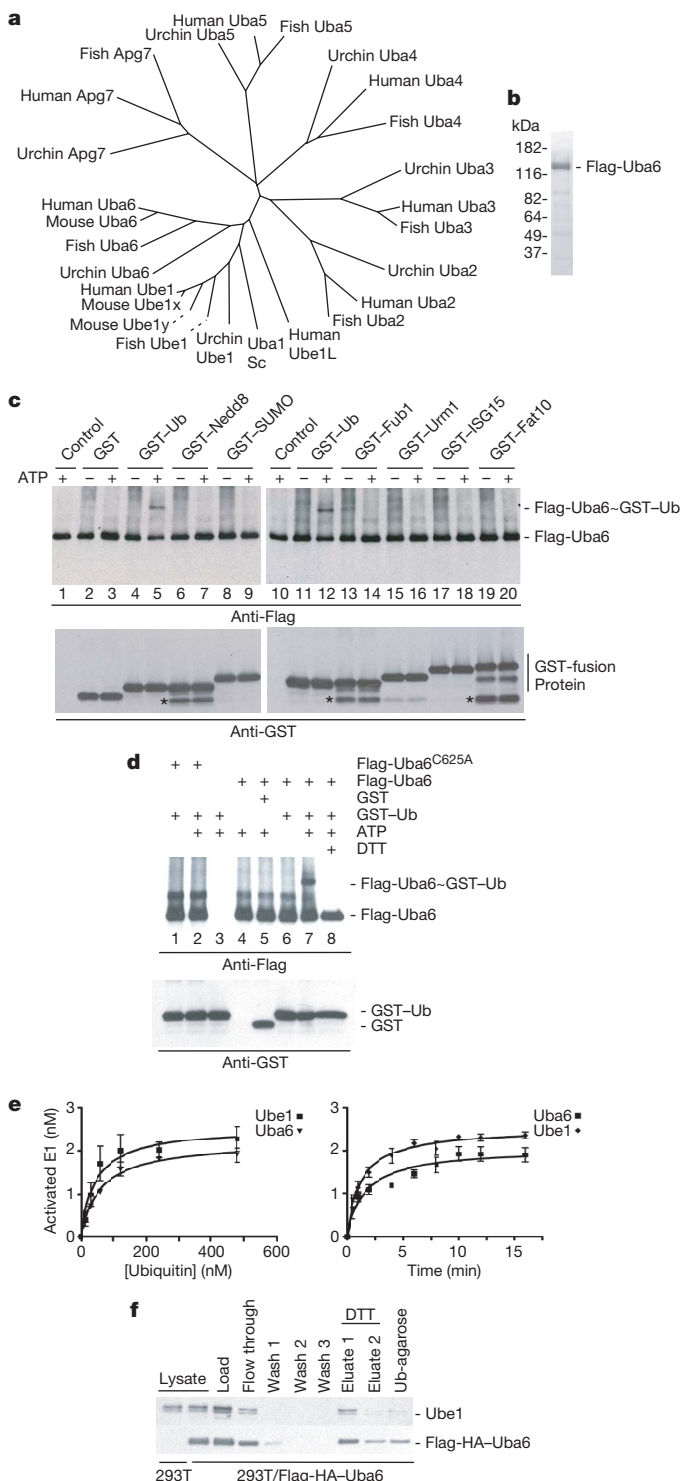


Figure 1 | Uba6 activates ubiquitin *in vitro*. **a**, Phylogenetic tree of ThiF-domain proteins in humans, zebrafish (*Danio rerio*) and sea urchin (*Strongylocentrotus purpuratus*). Also included are mouse Ube1x and Ube1y proteins as well as *Saccharomyces cerevisiae* Uba1p. **b**, Flag-Uba6 purified from insect cells (Coomassie blue staining). **c**, Flag-Uba6 activates GST-ubiquitin (GST-Ub). Activation was analysed by immunoblotting with anti-Flag or anti-GST. Asterisks indicate GST breakdown products. **d**, Ubiquitin activation by Uba6 requires ATP and Cys 625, and is sensitive to DTT. **e**, Kinetics of ubiquitin activation by Ube1 and Uba6 were determined as described in Methods. Error bars indicate s.d. for duplicate assays. **f**, E1 capture on ubiquitin-agarose was performed as described⁴⁻⁷, with 293T/Flag-HA-Uba6 extracts (pH 7.5).

to form dithiothreitol-resistant isopeptide linkages with lysine (Fig. 3b), which provided an independent assessment of E2 conjugation activity. The E2s tested fell into four classes (Fig. 3a, b): first, 14 E2s were charged efficiently by Ube1 but were not charged at all (for example the cell-cycle E2s Cdc34A and Cdc34B) or at much lower efficiency (for example E2A and E2B) by Uba6; second, nine E2s were charged by both Ube1 and Uba6 with similar efficiencies, including UbcH5 enzymes; third, five E2s were not charged by either Ube1 or Uba6; and last, an uncharacterized E2 (Gene ID 65264)²³, which we refer to as Use1 (Uba6-specific E2), was charged only by Uba6 (Fig. 3a, b, and Supplementary Fig. S5c). Thus, Uba6 and Ube1 have distinct specificities *in vitro*. Use1 orthologues are present in vertebrates and sea urchin (Supplementary Fig. S5b), and its mRNA is widely expressed in human tissues and cell lines (Supplementary Fig. S2).

Ufd8 recruit E2s to Nedd8 and SUMO-activating enzymes^{2,17,19,20,24}. We found that deletion of the Uba6 Ufd (Uba6^{ΔUfd}; Fig. 3c) abolished the charging of both Use1 and UbcH5D (Fig. 3c, and lane 7 in Fig. 3d), whereas replacement by the Ube1 Ufd (Uba6-Ufd^{Ube1}) resulted in charging of both Cdc34B and UbcH5D but loss of activity towards Use1 (Fig. 3d, lane 5). In contrast, the Ube1-Ufd^{Uba6} chimera charged UbcH5D but not Cdc34 or Use1 (Fig. 3d, lane 4). Thus, E1 Ufds

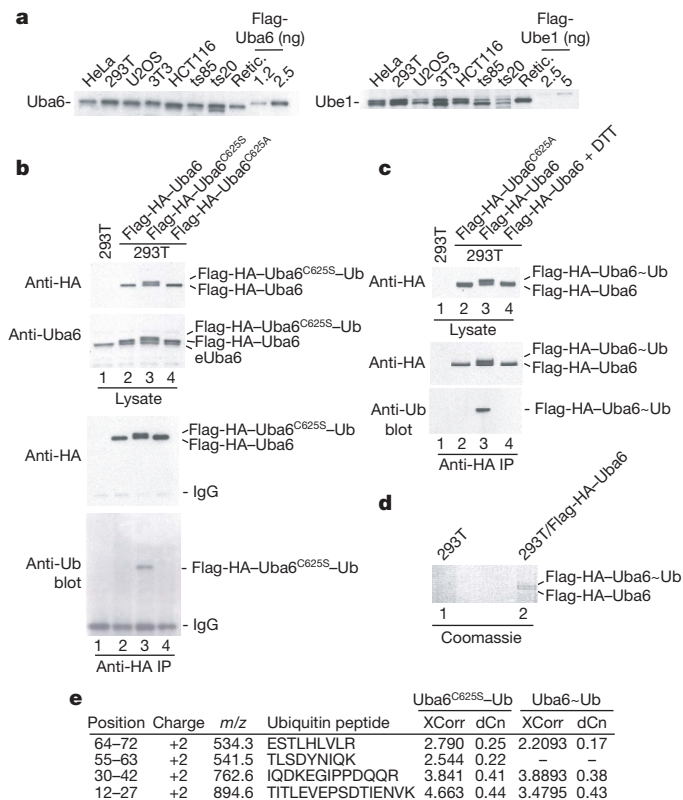


Figure 2 | Uba6 activates ubiquitin *in vivo*. **a**, E1 expression in cultured cells. Anti-Uba6 and anti-Ube1 were used to probe blots of extracts from the indicated cell lines, with recombinant Flag-E1s as controls. Retic., reticulocytes. **b**, Lysates (pH 7.5) from 293T/Flag-HA-Uba6 cells (wild-type, C625S or C625A; 20 μg) or anti-HA immune complexes from 0.2 mg of extract were separated on 4–12% Tris-glycine reducing gels and immunoblotted with anti-Uba6, anti-HA or anti-ubiquitin. IP, immunoprecipitation. eUba6, endogenous Uba6. **c**, Lysates (pH 4.5) from 293T/Flag-HA-Uba6 cells (wild-type or C625A; 20 μg) or anti-HA immune complexes from 0.2 mg of extract were separated on 4–12% Bis-Tris non-reducing gels before immunoblotting with anti-HA or anti-ubiquitin. **d**, Flag-HA-Uba6 was immunoprecipitated from 293T/Flag-HA-Uba6 cell lysates (pH 4.5, 2 mg) and separated on a 4–12% Bis-Tris non-reducing gel. **e**, Mass spectral analysis of ubiquitin-activated Uba6. The Flag-HA-Uba6^{C625S} protein used is shown in Supplementary Fig. S4c. XCorr, cross-correlation; dCn = (XCorr1 – XCorr2)/XCorr1.

contribute to selectivity in E2 recruitment, but interactions outside the Ufd not conserved in Ube1 probably contribute to the charging of Use1 by Uba6 (see Supplementary Information and Supplementary Fig. S7).

Because E1s and E3s interact with the same surface on E2s^{18,25}, the steady-state abundance of charged E2s *in vivo* is expected to depend on the rate of charging by E1s and on their rate of consumption by E3s and by spontaneous hydrolysis (Fig. 4a). Thus, E2 charging should be sensitive to the activity or abundance of E1s. To determine whether Uba6 and Ube1 differentially promote E2 charging, we developed a method for assessing the steady-state abundance of ubiquitin-charged E2s in cultured cells with the use of extracts generated at pH 3–5 (see Methods and Supplementary Fig. S5f). We found that most of Cdc34A and Cdc34B (which are both recognized by the Cdc34 antibody used) exist in more slowly migrating charged forms in HeLa cells that are abolished on the addition of reducing agent (Supplementary Fig. S5f). Charging of Use1 *in vivo* required its catalytic cysteine, and the charged form was sensitive to reducing agent, as determined with cells that stably express Flag-HA-Use1 or the C190A mutant (Fig. 4c). As with Cdc34, a substantial fraction of endogenous Use1 was charged *in vivo* (Fig. 4b, lane 1). We then examined the steady-state abundance of endogenous charged Use1 or Cdc34A and Cdc34B after the depletion of either Ube1 or Uba6 by using three independent short interfering RNAs (siRNAs). Ube1

depletion markedly decreased the charging of Cdc34A and Cdc34B but not that of Use1 (Fig. 4b, lanes 2–4). In contrast, Uba6 depletion greatly decreased the charging of Use1 but not that of Cdc34 (lanes 5–7). Simultaneous Uba6/Ube1 depletion did not further decrease charging beyond that seen with single siRNAs (lanes 8 and 9).

Thus, we have shown that Uba6 initiates an unexpected Ube1-independent pathway for the activation and charging of E2s such as Use1 with ubiquitin in vertebrates. Orthologues of Uba6 and Use1 are also present in the echinoderm sea urchin but seem to be absent from those arthropods, nematodes and fungi whose genomes have been fully sequenced. These findings challenge the view that activation and transfer of ubiquitin to all ubiquitin E2s is uniquely dependent on a single activating enzyme typified by Ube1 (refs 1, 8; Fig. 4d). Our results have several implications. First, the Uba6 and Ube1 cascades represent the only physiological example where a single UBL—ubiquitin itself—is used by two divergent E1s for transfer to a distinct set of E2s. A structural understanding of relevant E1/E2 pairs should help us to elaborate rules that govern E2 selectivity in UBL cascades. Further studies will also be needed to determine whether other E2s function exclusively with Uba6. Second, Uba6 and Ube1 are expressed together in many human tissues and in all the cell lines we have examined, indicating that the two ubiquitin activation pathways coexist in cells. Quantitative analysis suggests that Ube1 is more than tenfold more abundant than Uba6 in cell lines (Fig. 2a), which is consistent with the importance of Ube1 for the bulk of E2 charging in cells⁹. Nevertheless, our *in vivo* charging assay revealed specific roles for Uba6 and Ube1 in the charging of two

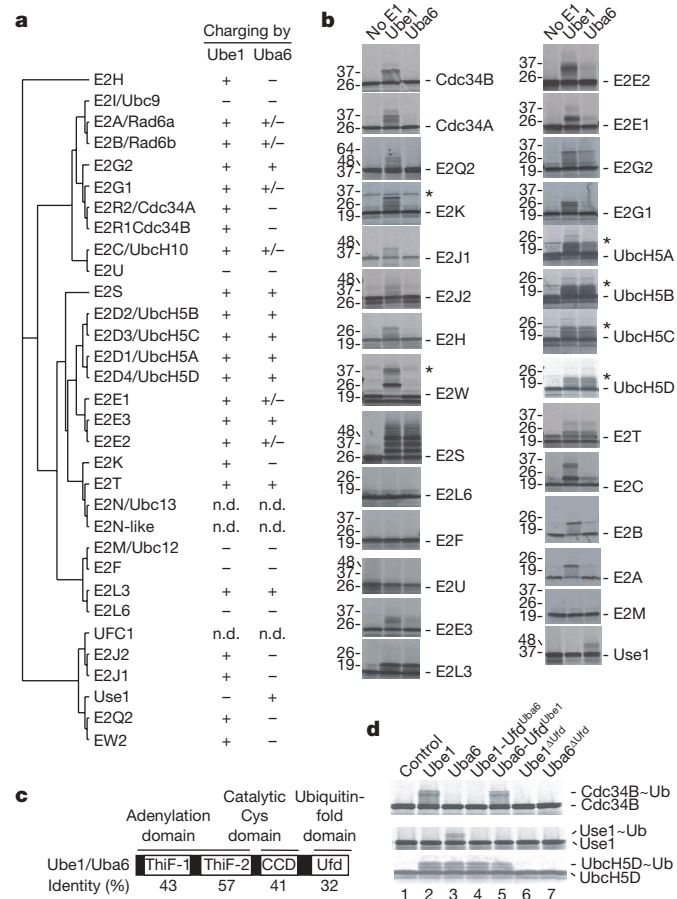


Figure 3 | Systematic analysis of E2-conjugating enzymes for targets of Uba6. **a**, E2-charging activity of ubiquitin E1s depicted on a phylogenetic tree of active E2s. **b**, Uba6 and Ube1 have distinct E2 charging activities *in vitro*. Assays employed ³⁵S-methionine-labelled E2s made in *Escherichia coli* S30 extracts, lysine-free ubiquitin and the indicated E1, as described in Methods. Asterisks indicate non-specific translation products. **c**, Sequence conservation of human Ube1 and Uba6. **d**, Charging of Cdc34B, UbcH5D and Use1 *in vitro* by chimaeric E1 proteins, examined with ³⁵S-methionine-labelled E2s.

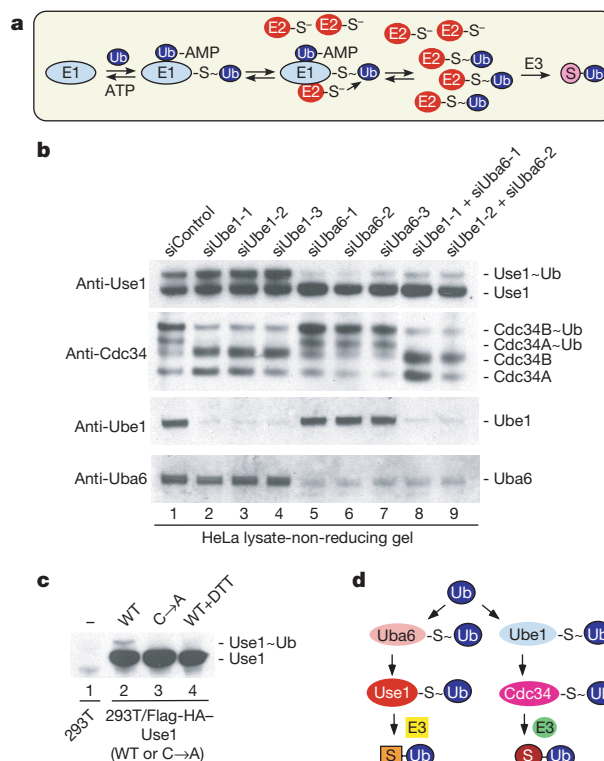


Figure 4 | Distinct requirements for charging of the ubiquitin conjugating enzymes Use1 and Cdc34 *in vivo*. **a**, Scheme depicting the mechanism of E2 charging in cells. E2s exist as a mixture of charged and uncharged forms, depending on how rapidly the charged form is generated and used. **b**, HeLa cells were transfected with the indicated siRNAs. After 72 h, cells were lysed at pH 4.5; proteins were then separated on a non-reducing 4–12% Bis-Tris gradient gel and immunoblotted. **c**, Use1 is charged in mammalian cells through its catalytic cysteine. 293T or 293T/Flag-HA-Use1 (wild-type (WT) or C190A) cells were lysed (pH 4.5) before separation on a non-reducing 4–12% Bis-Tris gradient gel and immunoblotting. In lane 4 the sample was pretreated for 5 min with 200 mM DTT. **d**, A model depicting two independent systems for activating and charging ubiquitin.

distinct ubiquitin E2s, namely Ube1 and Cdc34. Mouse cells that are temperature-sensitive for Ube1x (ref. 9) are frequently employed to determine whether a particular process involves the ubiquitin pathway. If the process in question is not blocked at a non-permissive temperature, it is considered to be independent of ubiquitin^{8,26–28}. Thus, it is conceivable that certain pathways that were previously thought to be independent of ubiquitin on this basis may nevertheless require ubiquitin by means of a Uba6-dependent pathway. These two E1 systems may also act either in parallel or in series to control biological processes. Although this is unlikely because of the high concentrations of ubiquitin in cells, it is also possible that Uba6 may activate other UBLs in specific cell types or in an inducible manner. Finally, a central conundrum concerns why an independent system for ubiquitin activation and E2 charging has emerged in vertebrates and echinoderms, when it might have been just as easy to adapt one of the several existing ubiquitin E2s to function together with Ube1 to target all critical substrates. One possibility is that Uba6 and Ube1 are differentially regulated by upstream signalling pathways to enhance flux through a specific conjugating pathway under particular circumstances. In this regard, we have found that whereas about 50% of Uba6 is activated with ubiquitin in proliferating 293T cells, Ube1 is fully activated (Supplementary Fig. S5h). This is consistent with the extent of charging of Cdc34 and Ube1 seen in parallel (Supplementary Fig. S5h) and suggests that flux through the Uba6–Ube1 pathway may be regulated. Identification of substrates and E3s that function specifically in the Uba6–Ube1 pathway (Fig. 4d) will be necessary to provide insight into this question. Such ubiquitin-dependent pathways would not be accessible through analysis of the conventional Ube1-activating enzyme alone.

METHODS SUMMARY

Ubiquitin activation assays contained 8 nM E1, 100 nM GST–UBL and 2 mM ATP in 50 mM Tris–HCl, 5 mM KCl, 5 mM MgCl₂ for 15 min at 30 °C (total volume 10 µl). Reactions were quenched with 2 × Laemmli sample buffer lacking reducing agent, and then subjected to non-reducing 4–12% Tris–glycine gel electrophoresis and immunoblotting. To examine E2 charging, the indicated E2 was translated *in vitro* with a bacterial S30 extract (Promega) in the presence of ³⁵S-methionine. E2 (1 µl) was incubated with 40 nM E1, 25 µM lysine-free ubiquitin (Boston Biochem), 2 mM ATP (15 min, 30 °C; 10 µl total volume) before 4–12% Tris–glycine gel electrophoresis followed by autoradiography.

For immunoprecipitation, cell extracts were generated in pH 7.5 lysis buffer (50 mM Tris–HCl pH 7.5, 150 mM NaCl, 0.5% Nonidet P40, protease inhibitors (Roche)) or pH 4.5 lysis buffer (50 mM MES pH 4.5, 150 mM NaCl, 0.2% Nonidet P40, protease inhibitors). Proteins were separated on non-reducing 4–12% Tris–glycine or 4–12% Bis–Tris gradient gels before blotting or mass spectrometry. To examine E2 charging *in vivo*, cells were lysed in 50 mM MES (pH 3–5 as indicated), 150 mM NaCl, 0.2% Nonidet P40 containing protease inhibitors and cleared by centrifugation. Extracts (10 µg) were subjected to non-reducing 4–12% Bis–Tris gel electrophoresis before immunoblotting. Where indicated, extracts were boiled (5 min) with 200 mM dithiothreitol (DTT). For RNA-mediated interference (RNAi), cells were transfected with Oligofectamine (Invitrogen). After 72 h, cells were lysed as described above. siRNA sequences were as follows: siUba6-1, 5′-CCTTGAAGAGAAGCCTGATGTAAA-3′; siUba6-2, 5′-ACACTGAAGTTATTGTACCGCATTT-3′; siUba6-3, 5′-GGGATCGATGG-ACCGTACATGGAAA-3′; siUbe1-1, 5′-GAGAAGCTGGCAAGCAGAAGTA-TT-3′; siUbe1-2, 5′-CCGACAGCTTGACTCCTACAAGAAT-3′; siUbe1-3, 5′-TCCTCAACTTGGCCCTGCCTTTCTT-3′.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 27 February; accepted 1 May 2007.

- Pickart, C. M. & Eddins, M. J. Ubiquitin: structures, functions, mechanisms. *Biochim. Biophys. Acta* **1695**, 55–72 (2004).
- Huang, D. T., Walden, H., Duda, D. & Schulman, B. A. Ubiquitin-like protein activation. *Oncogene* **23**, 1958–1971 (2004).
- Kerscher, O., Felberbaum, R. & Hochstrasser, M. Modification of proteins by ubiquitin and ubiquitin-like proteins. *Annu. Rev. Cell Dev. Biol.* **22**, 159–180 (2006).

- Ciechanover, A., Elias, S., Heller, H. & Hershko, A. 'Covalent affinity' purification of ubiquitin-activating enzyme. *J. Biol. Chem.* **257**, 2537–2542 (1982).
- Haas, A. L., Warms, J. V., Hershko, A. & Rose, I. A. Ubiquitin-activating enzyme. Mechanism and role in protein–ubiquitin conjugation. *J. Biol. Chem.* **257**, 2543–2548 (1982).
- Hershko, A., Heller, H., Elias, S. & Ciechanover, A. Components of ubiquitin–protein ligase system. Resolution, affinity purification, and role in protein breakdown. *J. Biol. Chem.* **258**, 8206–8214 (1983).
- Pickart, C. M. & Rose, I. A. Functional heterogeneity of ubiquitin carrier proteins. *J. Biol. Chem.* **260**, 1573–1581 (1985).
- Pickart, C. M. Back to the future with ubiquitin. *Cell* **116**, 181–190 (2004).
- Finley, D., Ciechanover, A. & Varshavsky, A. Thermolability of ubiquitin-activating enzyme from the mammalian cell cycle mutant ts85. *Cell* **37**, 43–55 (1984).
- Ciechanover, A., Finley, D. & Varshavsky, A. Ubiquitin dependence of selective protein degradation demonstrated in the mammalian cell cycle mutant ts85. *Cell* **37**, 57–66 (1984).
- McGrath, J. P., Jentsch, S. & Varshavsky, A. UBA1: an essential yeast gene encoding ubiquitin-activating enzyme. *EMBO J.* **10**, 227–236 (1991).
- Odorisio, T., Mahadevaiah, S. K., McCarrey, J. R. & Burgoyne, P. S. Transcriptional analysis of the candidate spermatogenesis gene Ubel and of the closely related Ube1x shows that they are coexpressed in spermatogonia and spermatids but are repressed in pachytene spermatocytes. *Dev. Biol.* **180**, 336–343 (1996).
- Lake, M. W., Wuebbens, M. M., Rajagopalan, K. V. & Schindelin, H. Mechanism of ubiquitin activation revealed by the structure of a bacterial MoeB–MoaD complex. *Nature* **414**, 325–329 (2001).
- Duda, D. M., Walden, H., Sfoudouris, J. & Schulman, B. A. Structural analysis of *Escherichia coli* ThiF. *J. Mol. Biol.* **349**, 774–786 (2005).
- Lehmann, C., Begley, T. P. & Ealick, S. E. Structure of the *Escherichia coli* ThiS–ThiF complex, a key component of the sulfur transfer system in thiamin biosynthesis. *Biochemistry* **45**, 11–19 (2006).
- Walden, H., Podgorski, M. S. & Schulman, B. A. Insights into the ubiquitin transfer cascade from the structure of the activating enzyme for NEDD8. *Nature* **422**, 330–334 (2003).
- Lois, L. M. & Lima, C. D. Structures of the SUMO E1 provide mechanistic insights into SUMO activation and E2 recruitment to E1. *EMBO J.* **24**, 439–451 (2005).
- Bencsath, K. P., Podgorski, M. S., Pagala, V. R., Slaughter, C. A. & Schulman, B. A. Identification of a multifunctional binding site on Ubc9p required for Smt3p conjugation. *J. Biol. Chem.* **277**, 47938–47945 (2002).
- Huang, D. T. *et al.* Structural basis for recruitment of Ubc12 by an E2 binding domain in NEDD8's E1. *Mol. Cell* **17**, 341–350 (2005).
- Huang, D. T. *et al.* Basis for a ubiquitin-like protein thioester switch toggling E1–E2 affinity. *Nature* **445**, 394–398 (2007).
- Haas, A. L. & Bright, P. M. The resolution and characterization of putative ubiquitin carrier protein isozymes from rabbit reticulocytes. *J. Biol. Chem.* **263**, 13258–13267 (1988).
- Komatsu, M. *et al.* A novel protein-conjugating system for Ufm1, a ubiquitin-fold modifier. *EMBO J.* **23**, 1977–1986 (2004).
- Gu, X. *et al.* Cloning and characterization of a gene encoding the human putative ubiquitin conjugating enzyme E2Z (UBE2Z). *Mol. Biol. Rep.* (in the press).
- Walden, H. *et al.* The structure of the APPBP1–UBA3–NEDD8–ATP complex reveals the basis for selective ubiquitin-like protein activation by an E1. *Mol. Cell* **12**, 1427–1437 (2003).
- Eletr, Z. M., Huang, D. T., Duda, D. M., Schulman, B. A. & Kuhlman, B. E2 conjugating enzymes must disengage from their E1 enzymes before E3-dependent ubiquitin and ubiquitin-like transfer. *Nat. Struct. Mol. Biol.* **12**, 933–934 (2005).
- Booth, J. W., Kim, M. K., Jankowski, A., Schreiber, A. D. & Grinstein, S. Contrasting requirements for ubiquitylation during Fc receptor-mediated endocytosis and phagocytosis. *EMBO J.* **21**, 251–258 (2002).
- Shringarpure, R., Grune, T., Mehlhase, J. & Davies, K. J. Ubiquitin conjugation is not required for the degradation of oxidized proteins by proteasome. *J. Biol. Chem.* **278**, 311–318 (2003).
- Chen, X. *et al.* N-acetylation and ubiquitin-independent proteasomal degradation of p21^{Cip1}. *Mol. Cell* **16**, 839–847 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Finley, B. Tansey, S. Elledge, J. Lou, R. Mulligan and B. Schulman for technical assistance, reagents and/or discussions, and B. Schulman and A. Sali for assistance with Modeller software. This work was supported by grants from the National Institutes of Health to J.W.H. and to S.P.G.

Author Information Sequences for human Uba6 and Ube1 have been deposited in the GenBank database under accession numbers EF623992 and EF623993, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.W.H. (wade_harper@hms.harvard.edu).

METHODS

Plasmids. Unless otherwise noted, open reading frames were cloned into pENTR/TOPO (Invitrogen) and transferred into the appropriate expression plasmid using *in vitro* recombination with Clonase (Invitrogen). The pHAGE-Flag-HA vector (puromycin resistant) places the open reading frame under control of the phosphoglycerate kinase promoter. Open reading frames for E2s were placed into vectors containing T7 promoters and an N-terminal His₆ tag. For expression of Use1 in bacteria, the Use1 open reading frame (NM_023079) was cloned into pENTR-2 (J.J. and J.W.H., unpublished observations) containing a Tobacco Etch Virus (TEV) protease cleavage site upstream of the open reading frame, and transferred into pDEST-15 (N-terminal GST tag from Invitrogen). The annotated open reading frame for Use1 (referred to as UBE2Z)²³ is incorrect, as determined by the size of the endogenous protein detected with anti-Use1 antibodies and by DNA sequence analysis. The actual open reading frame initiates 109 amino-acid residues before the annotated start site (Supplementary Fig. S5b). The sequences of all the genes examined are collected in Supplementary Table S1.

mRNA expression. Analysis of mRNA expression for Uba6, Ube1 and Use1 was performed with the Genomics Institute of the Novartis Research Foundation transcriptional profiling resource (<http://symatlas.gnf.org/SymAtlas/>)^{29,30}. This analysis employed GNF1H chips in combination with the MAS5 transcriptional profiling algorithm³⁰.

Phylogenetic trees. Trees and alignments were generated with ClustalW in conjunction with Treeview.

Antibodies. Anti-Uba6 and anti-Use1 antibodies were generated in rabbits by using a GST-Uba6 fusion protein (residues 869–1052) and GST-Use1, respectively, made in bacteria. Antibodies were affinity purified before use. The specificity of antibodies was determined by RNAi against Uba6 (Fig. 4) or Use1 (Supplementary Fig. S5g).

Protein expression and purification. For the production of proteins in insect cells, recombinant baculoviruses were used to infect *Spodoptera frugiperda* (Sf9) cells (40 h), and cleared cell extracts in lysis buffer (50 mM Tris-HCl pH 7.5, 0.5 mM DTT, 150 mM NaCl, 0.5% Nonidet P40, protease inhibitors) were bound to anti-Flag or glutathione-Sepharose beads. Washed beads were eluted with Flag peptide (500 µg ml⁻¹) or with glutathione (40 mM) and the protein was dialysed against 50 mM Tris-HCl pH 7.5, 0.5 mM DTT, 150 mM NaCl, 50% glycerol. Bacterial expression was performed in Rosetta/DE3 cells with induction by 0.4 mM isopropyl β-D-thiogalactoside for 3 h at 37 °C. Cells were disrupted in lysis buffer before purification with glutathione-Sepharose. All GST-UBL proteins were found to contain the appropriate C termini by mass spectrometry (see below) (Supplementary Fig. S3f).

Unless otherwise noted, expression of proteins in mammalian cells was accomplished by viral transduction with Gateway-compatible pHAGE (lenti-viral)-based vectors. Vectors were packaged in 293T cells and used to transduce the indicated cell line before selection with puromycin. For immunoprecipitation, extracts were generated in lysis buffer before incubation with the indicated antibodies. Proteins were separated on 4–12% Tris-glycine gradient gels (Invitrogen). To examine E2 charging, cells were lysed in buffer (50 mM MES, 150 mM NaCl, 0.2% Nonidet P40, protease inhibitors (pH 3–5 as indicated)) and cleared by centrifugation. Extract (10 µg) was subjected to non-reducing Novex 4–12% Bis-Tris gel electrophoresis before immunoblotting. When RNAi was employed, cells were transfected with the indicated siRNAs (33 nM) with Oligofectamine. After 72 h cells were lysed as described above, to examine E2 charging. The sequences of siRNAs are provided in Supplementary Table S1. To examine thioesters between ubiquitin and Flag-HA-Uba6, cells expressing Flag-HA-Uba6 (wild type or the C625A mutant) were lysed in 50 mM MES pH 4.5, 150 mM NaCl, 0.2% Nonidet P40 containing protease inhibitors, and extracts were subjected to immunoprecipitation with antibodies against HA-agarose. Immune complexes were washed with lysis buffer (pH 4.5), eluted with HA peptide, and then subjected to 4–12% Bis-Tris gel electrophoresis in the absence of reducing agent. In some cases, samples were treated with 200 mM DTT before electrophoresis, to reduce thioester bonds. Gels were transferred to poly(vinylidene difluoride) and probed with the indicated antibodies or, in some cases, bands were excised and subjected to mass spectrometry as described below.

In vitro ubiquitin activation and E2 charging assays. For ubiquitin activation shown in Fig. 1c, d, the indicated E1 (8 nM) was incubated with 100 nM GST-UBL and 2 mM ATP in reaction buffer (50 mM Tris-HCl pH 7.5, 5 mM KCl, 5 mM MgCl₂) for 15 min at 30 °C (total volume 10 µl); the reaction was quenched by the addition of 2 × Laemmli sample buffer lacking reducing agent, and immediately subjected to non-reducing 4–12% Tris-glycine gel electrophoresis and immunoblotting. To measure the kinetics of ubiquitin activation, reactions were performed with 8 nM E1 and 0.5 µM ubiquitin for the indicated duration at 30 °C and immunoblots were quantified with charge-coupled device detection of chemiluminescence. This was accomplished with an Alpha Innotech

FluorChem 8900 instrument. In experiments where the dependence on GST-ubiquitin concentration was determined, the GST-ubiquitin concentrations were 0, 7.5, 15, 30, 60, 120, 240 and 480 nM and the reaction time was 10 min. Mean and s.d. reaction rates for two independent experiments are shown. To examine E2 charging, the indicated E2 was translated *in vitro* with a bacterial S30 extract (Promega) in the presence of ³⁵S-methionine. Bacterial S30 extracts lack ubiquitin-activating and ubiquitin-conjugating enzymes, which could interfere with the assays. Radiolabelled E2 (1 µl) was incubated at 30 °C for 15 min with the indicated E1 at 40 nM and with 25 µM lysine-free ubiquitin and 2 mM ATP in reaction buffer (10 µl total volume). Reactions were analysed as described for ubiquitin activation. In some experiments, Use1 purified from bacteria was employed. Ube1 was released from GST-TEV-Use1 with TEV protease and the eluted Use1 was dialysed against 50 mM Tris-HCl pH 7.5, 0.5 mM DTT, 150 mM NaCl, 50% glycerol. Capture of Uba6 and Ube1 with ubiquitin-Sepharose was performed as described previously^{4,7} using extracts from cells stably expressing Flag-HA-Uba6.

Mass spectrometry. Unless otherwise stated, mass spectrometry was performed on peptides produced by in-gel trypsin treatment with a Thermo-Electron LTQ mass spectrometer. Searches were performed with Sequest. For the determination of C termini in GST-UBL fusion proteins, 2 µg of protein was excised from SDS-PAGE gels and digested with the protease indicated in Supplementary Fig. S3f. Digested peptides were then subjected to LC/MS/MS under conditions where more than 3,250 MS/MS scans were obtained for each LC/MS/MS run. The predicted C-terminal peptides as well as XCorr and dCn scores for the match with the predicted spectra. For analysis of peptides from Flag-HA-Uba6~Ub purified from mammalian cells, trypsin-treated samples were subjected to LC/MS/MS with an LTQ-Orbitrap instrument to identify peptides with high mass accuracy. Tandem mass spectra were first searched against the human genome with Sequest, identifying peptides against ubiquitin and Uba6 (see the text) but not other ubiquitin-like proteins. To rule out the presence of additional UBLs more effectively, a focused database containing protein sequences for human ubiquitin, Nedd8, SUMO-1, SUMO-2, SUMO-3, Fub1, Fat10, Urm1, Ufm1, Gdx, Isg15 and Apg12 was searched by using a 100-mass-unit filter (facilitated by the high mass accuracy afforded by Orbitrap detection). Peptides corresponding to ubiquitin alone were obtained (see Fig. 2), making it extremely unlikely that Uba6 efficiently charges UBLs in addition to ubiquitin.

29. Su, A. I. *et al.* Large-scale analysis of the human and mouse transcriptomes. *Proc. Natl Acad. Sci. USA* **99**, 4465–4470 (2002).

30. Su, A. I. *et al.* A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl Acad. Sci. USA* **101**, 6062–6067 (2004).

naturejobs

**JOBS OF
THE WEEK**

When I peruse rankings of the 'best movies of all time', the question inevitably pops into my head: are these truly the favourite movies of the 'experts' who picked them, or are they the ones they think most people regard as the best?

The same thoughts come to mind when it comes to rankings of universities, institutions, graduate programmes or postdoctoral experiences. What are the criteria? Who's judging them, and on what basis? We recently suggested that rankings of the world's universities often cement reputations, rather than offering useful, tailored advice (*Nature* **444**, 395; 2006). Good reputations beget good evaluations, and hence even better reputations.

Rankings are not going unchallenged, however. Last week, dozens of American liberal-arts colleges banded together to criticize the famous *U.S. News & World Report* annual US college rankings. At a meeting of 80 college presidents and 71 deans from the Annapolis Group of private colleges, participants rejected that part of the *U.S. News & World Report* system that asks presidents to gauge other schools' reputations, and most of them expressed their intent to withdraw from the ranking exercise completely. Instead, they pledged to work towards providing the equivalent information in a fairer form. The meeting follows a letter in May from the Education Conservancy, a non-profit organization based in Portland, Oregon, in which 12 college presidents asked colleges and universities not to provide information for the *U.S. News* list (*Nature* **447**, 514–515; 2007).

But although rankings are clearly less than perfect, they have a certain fascination. Perhaps it's the promise of guidance in a complicated world. Perhaps it's the irresistible desire to see where you fit in. Do I live in one of the top cities? Have I read the top novels? When it comes to higher education, though, should this instinctive response be resisted?

Get in touch and tell us what you think. Are college, graduate, and postdoc rankings useful, or are they a cottage industry that distorts more than it teaches and guides?

Gene Russo, acting editor, Naturejobs

CONTACTS

Acting Editor: Gene Russo

European Head Office, London

The Macmillan Building,
4 Crinan Street,
London N1 9XW, UK
Tel: +44 (0) 20 7843 4961
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Sales Manager:

Andy Douglas (4975)
e-mail: a.douglas@nature.com
**Business Development
Manager:**
Amelie Pequignot (4974)
e-mail: a.pequignot@nature.com

Natureevents:

Claudia Paulsen Young
(+44 (0) 20 7014 4015)
e-mail: c.paulsenyoung@nature.com

France/Switzerland/Belgium:

Muriel Lestringuez (4994)

Southwest UK/RoW:

Nils Moeller (4953)

Scandinavia/Spain/Portugal/Italy:

Evelina Rubio-Hakansson (4973)

Northeast UK/Ireland:

Matthew Ward (+44 (0) 20 7014 4059)

North Germany/The Netherlands:

Reya Silao (4970)

South Germany/Austria:

Hildi Rowland (+44 (0) 20 7014 4084)

Advertising Production Manager:

Stephen Russell
To send materials use London
address above.

Tel: +44 (0) 20 7843 4816

Fax: +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Jasmine Myer

US Head Office, New York

75 Varick Street, 9th Floor,
New York, NY 10013-1917
Tel: +1 800 989 7718
Fax: +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager:

Peter Bless

Japan Head Office, Tokyo

Chiyoda Building,
2-37 Ichigayatamachi,
Shinjuku-ku, Tokyo 162-0843
Tel: +81 3 3267 8751
Fax: +81 3 3267 8746

Asia-Pacific Sales Manager:

Ayako Watanabe
Tel: +81-3-3267-8765
e-mail: a.watanabe@natureasia.com

MOVERS

Thomas Mason, Director, Oak Ridge National Laboratory, Oak Ridge, Tennessee



2006-07: Associate Laboratory Director for Neutron Sciences, Oak Ridge National Laboratory (ORNL), Oak Ridge, Tennessee

2001-06: Associate Lab Director for the Spallation Neutron Source, ORNL

1998-2001: Director of the Experimental Facilities Division of the Spallation Neutron Source, ORNL

With a physicist father and biochemist mother, it never occurred to Thomas Mason to be anything other than a scientist. Nevertheless, a big-picture, wide-ranging view has propelled his career since its early years.

The native Canadian made his first crucial career decision as a physics PhD student at McMaster University in Hamilton, Ontario. Choosing Malcolm Collins, who worked mostly at large facilities, as his adviser directed Mason towards high-profile physics laboratories. He did his PhD work on neutron scattering at Chalk River, a Canadian national laboratory. "My decision meant I got involved with a much larger range of projects than I otherwise would have," says Mason. He published 17 papers from his PhD work — more than half of which weren't part of his thesis.

Mason then did a postdoc at Bell Labs in Murray Hill, New Jersey, at the tail end of the famed lab's heyday. "It was an environment that made you better because of the peers egging you on," he says.

Mason thought his career moves were over when he accepted an associate professorship at the University of Toronto, Ontario. But five years later, Oak Ridge National Laboratory (ORNL) tapped Mason to be the scientific director on the not-yet-built Spallation Neutron Source (SNS), the laboratory's accelerator-based neutron source.

Mason surprised many by leaving the University of Toronto a week after he received tenure. "I didn't realize at the time what a crazy idea it was to leave tenure and join a new, risky project," he says, noting that the cancellation of the superconducting supercollider made it somewhat difficult to recruit scientists for the SNS, another big-budget government project.

Former ORNL director William Madia picked Mason to head the SNS because, he says, all the other international candidates were focused only on their own discipline. "Thom was the only one who saw what it could do in all fields of science," Madia says. He watched as Mason spent \$1 million a day to bring the project in on time and on budget — and he began grooming Mason for bigger responsibilities.

The challenge now, says Madia, is to produce a scientific programme that can maximize the return on high-profile ORNL investments like the SNS and the recently constructed High Flux Isotope Reactor. Mason becomes ORNL's director on 1 July to do just that. "We've built the Maserati, now Thom's going to drive it," says Madia. Virginia Gewin

NETWORKS & SUPPORT

Making contacts online

Most people manage their contacts with the aid of a personal digital assistant (PDA), cell phone or computer. Now a new generation of online tools are helping professionals get even better connected.

Free networking platforms are proliferating on the Internet. Just as young people use websites such as MySpace to find friends and share photos and music, business people use online networks such as LinkedIn and Ryze to make new contacts and discover new opportunities.

Physicians and life-sciences researchers can now do the same through the online biomedical community PrometeoNetwork (www.PrometeoNetwork.com) via a platform provided by the company Within3 (www.Within3.com). Within3 provides the ability to find collaborators, refer patients and find new professional opportunities, in addition to making it easier to manage existing personal contacts.

PrometeoNetwork hosts several subgroups, including, for example, 'stem-cell researchers', 'Italian scientists' or 'oncologists'. Members are invited to join such groups by colleagues. PrometeoNetwork and Within3 organize promotions, social gatherings and participation at conferences. PrometeoNetwork

is also working to raise funds for fellowships and to help members find opportunities to work and study abroad.

Members build a profile of their background and interests, and can associate their publications and clinical trials with their profile using a tool that interacts directly with the relevant databases. The community grows as members invite their peers to join (administrators routinely evaluate new requests for membership). Using the search engine, members can find and filter information that exists in the network and in other databases, discover new colleagues, stay current on the work of their peers, communicate with other physicians and scientists, and build their personal networks. Through privacy settings, members can control who has access to their information.

Face-to-face contact at conferences will always be important to scientists. But the new technologies are ways to augment traditional networking. The chances of finding a key collaborator can be dramatically improved — without ever leaving the lab bench.

Fabio Fais is a medical biotechnologist and a manager at PrometeoNetwork.

POSTDOC JOURNAL

The final push

Preparing for a hiatus in your work is hard enough when you can plan it around set dates. When you are trying to plan it around biology, all bets are off.

There is a palpable nervousness in my lab these days, and it isn't emanating from me. Some of the unspoken fear is that my waters will break in the lab, humiliating everyone. Oddly, this isn't a fear I share. The other postdoc made some half-hearted joke about maybe getting some emergency medical training in case I needed 'help' one day in the lab. Again, not a fear I share as I will not deliver the baby in the lab. All these concerns felt silly to me, and so far removed from what I am really worrying about.

My worries centre around the void I will inevitably leave. Although they're not that critical, who will be delegated to do my daily lab duties? My biggest concern, however, is my absence from the field nursery. I am planting more than 24,000 plants for a phenotyping experiment I might never see. If the baby arrives early, I might be able to participate, but there is no guarantee that everything will go as planned. So in preparing to miss out on my largest field experiment ever, I am trying to iron out all the organizational issues now. Ironically, no one else in the lab is stressed out about this at all: everyone is ready to lend a hand.

Moir Sheehan is a postdoc in plant breeding and genetics at Cornell University.

The inside track from academia and industry

Taking on that first faculty job

You know the science, but there's plenty else to consider.



David Burgess

Scientists landing their first faculty job at a research university are well prepared to pursue their science. But many are not prepared for all the ancillary problems involved in becoming a faculty member and eventually being successful in getting tenure. Those from under-represented minorities face additional hurdles. A fortunate few will have had good advice from a mentor or programme, but for the rest I offer some practical advice that may help in the quest for tenure.

Highly recommended is a written plan with realistic annual goals for the first three or four years. Write down your own measures for success (hiring staff, getting funded, generating data for papers, publishing). After a year, review your plan and modify it as needed.

By speaking with those who have recently gone through the tenure process in your department, you can learn about numbers of papers, amount of grant funding, level of service, national visibility, presentations at meetings, and so on, that the department and school expect for the award of tenure. You might share your plan with one of them, perhaps suggesting that they become your mentor, to get some frank feedback.

In your own lab you will set the tone for those you hire as technicians and train as postdocs and grad students. From work hours to radio stations, from lab meetings to socializing, your initial decisions are crucial because once the tone is set it's not likely to change.

Managing time is the most critical element to success as an assistant professor. Those with teaching responsibilities can expect to spend — believe it or not — about 10 hours of

preparation time for each hour lecture the first time they teach the course.

Serving on a departmental committee is a double-edged sword. Too much, or the wrong kind, becomes a time sink that hinders your progress. Too little and your colleagues won't consider you one of the team. I recommend volunteering to be on the graduate admissions committee or the seminar committee (or the equivalent at your university). The seminar committee provides a way to document the national visibility component in your tenure file.

Network with your friends who have recently been hired and arrange for reciprocal invited seminar visits. Some national service enhances both your service record and your research. Accept invitations to be an *ad hoc* grants reviewer for the NIH,

the NSF or private funding agencies. You may join a department where there are political camps and feuding faculty. Stay low; try not to

be enticed to join one camp or another. Keep your opinions close at first until you learn more about the faculty and their issues.

For under-represented minorities and women there are extra problems. New female assistant professors tend to be at the age when the biological clock is ticking. Ask questions about whether the tenure clock is halted for birth or adoption. Ask about day care.

Women need to look at the department to see whether women got tenure and were then promoted to full professor. How thick is the glass ceiling for women in the department? What efforts have been made to address the local career path for women in your institution?

Minority faculty are so rare

at research universities that they face both expected and unexpected challenges and barriers not faced by their white colleagues. Don't be surprised if, as a new minority faculty member, you're called by the dean, the provost or the president asking you to serve on an important task force or committee. My advice: learn to say "no" politely. Tell them that you'll be glad to be on such a committee after you get tenure; it's best to focus your energies on your research at this stage of your career. Even if you offend them, don't worry — the half lives of administrators are so short that they will probably be gone by the time you come up for tenure.

Minority science faculty are also sought after by all the minority students for advice, friendship and mentoring. Try to limit these richly deserved opportunities as best you can — again until after tenure, when your lab is established and you have more freedom. Minority faculty often get invited to serve on more national committees than others; pick those that will enhance your research career.

Women and minorities also have to contend with being judged by different standards from their white male assistant professor colleagues. A white male is expected to be assertive or aggressive, but a minority person or a woman who acts in the same way is regarded by some as being uppity, rude or too aggressive. Recognize that some of your colleagues have never been around successful and brilliant women or minorities and will act accordingly.

Finally, keep focused, remember your family and make time for them. And, if at all possible, enjoy yourself. ■

David Burgess is a professor of biology at Boston College, a past president of the Society for the Advancement of Chicanos and Native Americans in Science, and a long-time mentor.

"Learn to say 'no' politely. Tell them that you'll be glad to be on the committee after you get tenure."